

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Ciências Biológicas
(Bioquímica)

VINÍCIUS FERREIRA DA PAIXÃO

**Anotação e caracterização de novos transcritos
expressos no adenocarcinoma de pâncreas:
formas alternativas de *splicing* e RNAs não-
codificadores longos associados a fenótipos
tumoriais e clínicos**

Versão original da Tese defendida

São Paulo

Data do Depósito na SPG:

22/08/2019

VINICIUS FERREIRA DA PAIXÃO

**Anotação e caracterização de novos transcritos
expressos no adenocarcinoma de pâncreas:
formas alternativas de *splicing* e RNAs não-
codificadores longos associados a fenótipos
tumorais e clínicos**

*Tese apresentada ao Instituto de Química
da Universidade de São Paulo para
obtenção do Título de Doutor em
Ciências Biológicas (Bioquímica)*

Orientador: Prof. Dr. Eduardo Moraes Rego Reis

São Paulo

2019

Dedico a Deus na pessoa de meus pais, Luzia e Ricardo, e meu irmão, Bráulio; exemplares da substância sem moléculas mais preciosa do universo: o amor.

A meu avô Antônio Gomes Ferreira, in memoriam, a todos meus familiares, amigos e a Júlio César Gomes de Sousa Filho, in memoriam.

AGRADECIMENTOS

Primeiramente agradeço a Deus pelo dom da vida e pela oportunidade de tentar aprender um pouco mais sobre o complexo e refinado universo molecular da vida. Agradeço imensamente a minha família: meus pais Luzia e Ricardo e meu irmão Bráulio, os quais jamais mediram esforços para o meu crescimento e seguimento na vida acadêmica, nunca foi fácil. Agradeço a imensa compreensão dos meus pais, que abriram mão de conforto, horas de descanso e segurança para que eu pudesse estar onde estive. Pai e mãe, acreditem quis fazer “*valer a pena*” dos seus calos dos pés, da técnica em enfermagem dedicada, e dos seus calos das mãos, do pedreiro talentoso. Obrigado pelas rugas e lágrimas de preocupação com meu presente e futuro, rezando para que eu não adoecesse pois sem seguridade só pudemos recorrer mesmo a Deus. Agradeço meu irmão e grande amigo Bráulio, pelo companheirismo, cuidado e constantes ajudas durante praticamente todo o tempo do doutorado.

Agradeço a meus familiares que sempre se dispuseram em ajudar naquilo que fosse possível. Agradeço minha avó Maria pelas orações e cuidado mesmo a distância e meus tios pela preocupação em consolar meus pais pela *síndrome do ninho vazio*. Agradeço a cada um pela compreensão sobre as minhas ausências em tantos momentos difíceis e nos momentos mais comuns, de felicidade. Agradeço a cada um sobretudo, pela insistência em me procurarem nos momentos nos quais eu estava absorto em meus problemas, sempre com uma palavra para me dar ânimo e boas risadas.

Agradeço àqueles meus colegas estudantes, meus professores do ensino básico, ensino fundamental, ensino médio (a esses agradeço nominalmente a professora Elisa por mostrar como a bioquímica era algo apaixonante), professores do ensino superior e os amigos pesquisadores (pós-graduandos): a todos vocês obrigado pela fidelidade ao sonho altruísta de gerar conhecimento para que a vida e a sociedade sejam cada vez melhores, vocês são heróis.

Agradeço imensamente ao Dr. Prof. e orientador Eduardo Moraes Rego Reis pela oportunidade que me concedeu de integrar sua equipe de pesquisa, a imensa atenção e solicitude na rotina do laboratório, os momentos de críticas construtivas, a dedicação e paciência. Além do estímulo para praticar esportes!

Agradeço a melhor equipe de pesquisa que já existiu: Bianca Dazzani, Bruno Cruz, Diogo Pessoa, Diogo Pellegrina, Ester Cabrita, Thalita Bueno, Luíza Scalabrini, Lutero, Dimitrius, Damian e todos os outros do time LAPIC. Aos colaboradores da iniciação científica, com quem tive a oportunidade de aprender: Nathalia Lainetti, Victória, Ricardo Chiong, Juliana.

Agradeço as mulheres mais admiráveis que encontrei no mundo da ciência: Layla Farage Martins e Bianca Dazzani (sim, menciono pela segunda vez) que com muita competência e habilidade tornam as hipóteses testáveis e a convivência nos laboratórios momentos felizes.

Agradeço imensamente a Canuta por sempre manter o ambiente agradável para se fazer ciência, por sempre me receber com muita alegria no corredor do bloco 12 independentemente do quão frio estivesse.

Agradeço aos professores que compartilharam a sala de aula comigo e outros monitores PAE: Maria Terêsa Machini, Graziella Eliza Ronsein, Roberto Kopke Salinas, Nicolas Carlos Hoch e todos os demais, sempre me mostrando como o ensino é algo sublime e transformador.

Agradeço aqui e peço perdão a todos amigos que (por falta de tempo, espaço e imperdoável falta de memória) não foram aqui citados, vocês sempre me deram a coragem e muitas vezes foram a força motriz para que eu realizasse muitas coisas.

Agradeço profundamente ao meu amigo de infância Pékos, que me acolheu em São Paulo e me ajudou a me estabelecer nessa *urbe et orbi*.

Agradeço aos melhores mineiros que resolveram enfrentar São Paulo num apartamento na Bela Cintra, Fahyme e Raphael Messias, com quem partilhei muitos momentos importantes, as conversas no fim do dia sobre ciência, religião, política e mais uma gama de pautas aleatórias que não caberia em nenhum livro. Guardo com muito carinho essas memórias e digo que: sem vocês a vida nessa cidade não teria tido a menor graça.

Agradeço imensamente a Companhia de Jesus, nominalmente ao Ir. Ubiratan Costa, *sj* e Padre Jonas Elias Caprini, *sj* por me ajudar a buscar e viver o *magis* na conclusão do doutorado, para *em tudo amar e servir*, para viver consolado. Bem como meus mais sinceros agradecimentos a todos do Centro Magis *Anchietanum*, pela acolhida, compreensão e solicitude.

Agradeço imensamente a ilustríssima instituição do Instituto de Química da Universidade de São Paulo (IQ-USP), que me acolheu através do programa de Pós-Graduação em Ciências Biológicas (Bioquímica) onde foi realizado este doutorado, bem como as agências de fomento que financiaram projetos nos quais estive envolvido e a bolsa, FAPESP e CNPq.

Enfim, a todos meus mais sinceros agradecimentos, esse trabalho nunca foi meu, mas sempre resultado de um coletivo amoroso.

“Porque onde está o teu tesouro, lá também está teu coração.”

(Mateus 6, 21)

“Três ideias profundamente desestabilizadoras ricochetearam por todo século XX e se dividiram em três partes desiguais: o átomo, o byte e o gene”

(BAUER, Martin W.)

RESUMO

(Paixão, V. F.) **Anotação e caracterização de novos transcritos expressos no adenocarcinoma de pâncreas: formas alternativas de splicing e RNAs não-codificadores longos associados a fenótipos tumorais e clínicos.** 2019. Número de páginas do trabalho (145p). Tese (Doutorado) - Programa de Pós-Graduação em 2019. Instituto de Química, Universidade de São Paulo, São Paulo.

O câncer de pâncreas (PC) é uma das doenças mais devastadoras com o pior resultado de sobrevivência quando comparada com outros tipos de câncer. É apontado como o décimo câncer mais comum e a quarta *causa mortis* por câncer, projetando-se para ser a segunda até 2030 nos Estados Unidos e na Alemanha. O PC constitui um conjunto heterogêneo de tumores, o adenocarcinoma ductal (PDAC) constitui o tipo mais frequente da neoplasia (80%). A prevenção, detecção precoce e tratamento enfrentam sérios problemas e é urgente a identificação de novos marcadores para diagnóstico, prognóstico e estratégias terapêuticas da doença. O objetivo desse trabalho foi realizar uma análise com alta-resolução do transcriptoma do PDAC) para identificar novos RNAs não codificadores, variantes de *splicing* e alterações transcricionais associados à malignidade. A metodologia empregada constituiu-se de gerar bibliotecas de RNA total a partir de 14 amostras cirúrgicas pareadas de tecido tumoral (PDAC) e tecido não-tumoral adjacente para sequenciamento NGS. A partir dos dados de RNAseq foi realizada a montagem do transcriptoma utilizando-se a referência do catálogo GENCODE para classificar os transcritos. Seguiram-se análises subsequentes de avaliação de características dos transcritos: estruturais, marcas regulatórias, potencial codificador, detecção em banco de dados independente (*miTranscriptome*, miT). Em segundo lugar foram realizadas análises de expressão diferencial, análise de sobrevida (utilizando banco de dados públicos) para seleção de transcritos potencialmente relevantes no PDAC. Foram geradas redes de co-expressão gênicas com enriquecimento de funções biológicas. Uma série de validações foram realizadas por RT-PCR de transcritos novos intergênicos e novas formas de *splicing* reconstruídas e selecionadas. RT-qPCR foi utilizada para validação da expressão aberrante de lncRNAs em PDAC, silenciamentos por siRNA e subsequentes ensaios de proliferação, migração e invasão. Foi avaliada *in vivo* o envolvimento com crescimento tumoral por ensaio xenográfico, e avaliação da implicação em funções biológicas mais específicas, como envolvimento em reparo de DNA por ensaio cometa alcalino e avaliação de enriquecimento em tumoresferas. A montagem reconstruiu 90.522 transcritos, dos quais 41.341 são anotados no GENCODE e 6.710 transcritos são novos não anotados no GENCODE (classificado como *splicing variant*, *intergenic RNA*, *antisense RNA*). Desses foram validados 6 novos lincRNAs com expressão aumentada em PDAC, dois deles (TCONS00085964 e TCONS00036574) tem a expressão correlacionada com alterações na sobrevida. Foram validadas também novas formas de *splicing* de MMP14, CAPN8, LIF e OCT3 com expressão aumentada em PDAC. 7 lncRNAs, anotados no GENCODE, com expressão aumentada em PDAC, desses foram implicados com fenótipo tumoral de migração, invasão e proliferação: LINC01559; LINC01133, CCAT1 e UCA1. Desses LINC01559 e

CCAT1 demonstrara regular a expressão de enzimas de O-glicosilação (GALNT3 e B3GNT3) envolvidas na manutenção de células tronco-tumorais em PDAC. O lncRNA UCA1 está envolvido com reparo de DNA e envolvido com a progressão tumoral *in vivo*. Conclui-se que a abordagem por amostras pareadas a partir de RNAseq de bibliotecas de RNA total gera um catálogo de transcritos, um transcriptoma mais apurado e que lncRNAs estão implicados em PDAC, como LINC01559 e UCA1 tem potencial translacional e abrem caminhos para o desenvolvimento de estratégias translacionais.

Palavras-chave: Transcriptoma, xenotumor, PDAC, lncRNA, *alternative splicing*, UCA1.

ABSTRACT

(Paixão, V.F.) **Determination of relevant transcripts in ductal adenocarcinoma of pancreas: the transcriptional landscape of the long non-coding RNAs and protein-encoding RNAs revealed by the total RNAseq analysis of pancreatic adenocarcinoma.** 2019. (145p.). PhD Thesis - Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Pancreatic cancer (PC) is the deadliest malignancy, one of the most devastating diseases with the worst survival outcome compared to any cancer. It is **fourth** as the tenth most common cancer and the fourth leading cause of cancer in the United States and Germany. It is projected to be the second leading cause of cancer death in a decade. PC is a heterogeneous set of tumors with high aggressiveness and mortality. Of these tumors, ductal adenocarcinoma (PDAC) is the most frequent type of cancer (80%). Prevention, early detection and treatment face serious problems, and the identification of new markers for diagnosis, prognosis and therapeutic strategies of the disease is urgent. The aim of this study was to perform a high-resolution analysis of pancreatic adenocarcinoma (PDAC) transcriptome to identify new noncoding RNAs, splicing variants, and transcriptional changes associated with malignancy in pancreatic ductal adenocarcinoma. The methodology employed consisted of generating total RNA libraries from 14 paired surgical samples of tumor tissue (PDAC) and adjacent non-tumor tissue for NGS sequencing. From the RNAseq data, the transcriptome assembly was performed using the GENCODE catalog reference to classify the transcripts. Subsequent analyzes of evaluation of transcript characteristics were followed: structural, regulatory markers, potential encoder, independent database detection (miTranscriptome, miT). Secondly, differential expression analysis, survival analysis (using public database) were performed to select potentially relevant transcripts in the PDAC. Genic co-expression networks with biological function enrichment were generated. A series of validations were performed by RT-PCR of new intergenic transcripts and new forms of reconstructed and selected splicing. RT-qPCR was used for validation of aberrant expression of lncRNAs in PDAC, siRNA silencing and subsequent proliferation, migration and invasion assays. Involvement with tumor growth was evaluated by *in vivo* xenograph assay. Biological function tests to determine an involvement in DNA repair was evaluated by alkaline comet assay and was performed an evaluation of tumorsphere expression enrichment. The assembly reconstructed a total of 90,522 transcripts, of which 41,341 are noted in GENCODE. 6,710 transcripts are new (splicing variant, intergenic RNA, antisense RNA) not noted in the GENCODE reference. Of these 6 new PDAC-enhanced lincRNAs were validated, two of them (TCONS00085964 and TCONS00036574) correlated with changes in survival. New forms of splicing of MMP14, CAPN8, LIF and OCT3 with increased expression in PDAC were also validated. 7 lncRNAs noted with increased expression in PDAC were implicated with tumor migration, invasion and proliferation phenotype: LINC01559; LINC01133, LINC01614; CCAT1; LINC02577; LINC00920 and UCA1. Of these LINC01559 has been shown to regulate the expression of O-glycosylation enzymes (GALNT3 and B3GNT3) involved in maintaining CSCs in PDAC. It has been identified that UCA1 is

involved with DNA repair and involved with tumor progression *in vivo*. It is concluded that the approach of sampling RNAseq from total RNA libraries generates a transcript catalog, a more accurate transcriptome and that lncRNAs are implicated in PDAC, such as LINC01559 and UCA1 that have potential prognostic use and open ways for development of translational strategies.

Keywords: Transcriptome, xenotumor, PDAC, lncRNA, alternative splicing, UCA1.

Lista de abreviaturas e siglas

°C – graus Celsius

µg – micrograma

µL – microlitro

µM – micromolar

AEC – Antes da Era Cristã

Anvisa – Agência Nacional de Vigilância Sanitária

AS – *Splicing* alternativo, do inglês *alternative splicing*

BC – Câncer de mama, do inglês *breast cancer*

bp – pares de base, do inglês *base pairs*

CA 19-9 – Antígeno Carcinogênico 19-9, do inglês *cancer antigen 19-9*

CAF – Fibroblastos associados ao câncer, do inglês *cancer associated fibroblasts*

cDNA – DNA complementar, do inglês *complementary DNA*

CEA – Antígeno carcinoembrionário, do inglês *carcinoembryonic antigen*

ceRNA – RNA endógeno competitivo, do inglês *competitive endogenous RNA*

CRC – Câncer Colorretal, do inglês *colorectal cancer*

CSC – células tronco tumorais, ou *cancer stem cells*

CT – Tomografia computadorizada, do inglês *Computed tomography*

DEC – Depois da Era Cristã

DMEM – Meio Eagle modificado de Dulbecco, do inglês *Dulbecco's modified Eagle's medium*

DNA – Ácido desoxirribonucleico, do inglês *deoxyribonucleic acid*

dsDNA – DNA dupla fita, do inglês *double strand DNA*

ESCC – câncer epidermóide de esôfago, do inglês *esophageal squamous cell carcinoma*

g – grama

gDNA – DNA genômico, do inglês *genomic DNA*

HCC – Carcinoma hepatocelular, do inglês *hepatocellular carcinoma*

HGP – Projeto Genoma Humano, do inglês *Human Genome Project*

HR – Recombinação homóloga, do inglês *homologous recombination*

HR – Taxa de risco, do inglês *hazard ratio*

IPMN – Neoplasia intraductal papilar mucinosa, do inglês *Intraductal papillary mucinous neoplasm*.

L – Litro

lncRNA – RNA longo não-codificante, do inglês *long non-coding RNA*

M – Mol/L molar

min – minutos

miRNA – microRNA

mL – mililitro

NCD – Doenças não transmissíveis, do inglês *noncommunicable disease*

ncRNA – RNA não-codificante, do inglês *non-coding RNA*

nm – nanômetro

NMF – Fatoração de matrizes não-negativas, do inglês *non-negative matrix factorization*

NMF – Fatoração em Matrizes Não Negativas, do inglês *Non-negative matrix factorization*

no RT – reação controle sem transcriptase reversa, do inglês *no reverse transcriptase control*

NSCLC – Câncer de pulmão de células não pequenas, do inglês *non small cell lung cancer*

nt – nucleotídeo

NTC – reação de controle sem cDNA molde, do inglês *non template control*

PanIN – Neoplasia intraepitelial pancreática, do inglês *Pancreatic intraepithelial neoplasia*.

PC – Câncer de Pâncreas, do inglês *Pancreatic Cancer*

PCA – Análise em Componentes Principais, do inglês *Principal Component Analysis*

PCR – reação em cadeia da polimerase, do inglês *Polimerase Chain Reaction*

PCSC – Célula tronco-tumorais pancreáticas, do inglês *pancreatic cancer stem cell*

PDAC – Adenocarcinoma ductal de pâncreas, *Pancreatic Ductal Adenocarcinoma*

PDO – Organóide derivado de paciente, do inglês *patient-derived organoid*

PDX – xenotumores derivados de pacientes, do inglês *Patient derived xenografts*

PSCs – células pancreáticas estreladas, do inglês *pancreatic stellate cells*

qPCR/RT-qPCR – reação da transcriptase reversa, seguida de reação em cadeia da polimerase, do inglês *reverse transcription polymerase chain reaction quantitative real time*

RCEX – Rede de Co-expressão

RNA – Ácido ribonucleico, do inglês *ribonucleic acid*

RPMI – Meio Roswell Park Memorial Institute, do inglês *Roswell Park Memorial Institute médium*

RT – Transcriptase reversa, do inglês *reverse transcription*

ssDNA – DNA fita simples, do inglês *single strand DNA*

TME – Microambiente tumoral, do inglês *tumour microenvironment*

TSAs – Antígenos específicos de Tumores, do inglês *tumour-specific antigens*

ng – nanograma.

Lista de Figuras

Figura 1 – Evolução das lesões precursoras para o PDAC.	26
Figura 2 – Árvore filo-transcriptômica de subtipos de expressão molecular de PDAC.	36
Figura 3 – <i>Splicing</i> constitutivos e alternativos.	41
Figura 4 – <i>Hallmarks</i> do <i>splicing</i> alternativo em câncer.	42
Figura 5 – Classificação de ncRNAs. Uma proposta de divisão em classes e subclasses em eucariotos.	44
Figura 6 – Modelos de funções atribuídas a lncRNAs.	46
Figura 7 – lncRNAs associados as <i>hallmarks</i> do câncer.	48
Figura 8 – Fluxograma geral da rotina empregada na análise computacional dos dados de RNAseq de tumores de pâncreas.	60
Figura 9 – Detecção dos transcritos não anotados no GENCODE detectados na montagem final dos nossos dados de RNAseq de PDAC com o catálogo miTranscriptome (miT).	77
Figura 10 – Novos transcritos intergênicos selecionados e validados.	79
Figura 11 – <i>Heatmap</i> das expressões relativas dos novos transcritos intergênicos validados em linhagens celulares de PDAC por RT-qPCR.	80
Figura 12 – Validação da expressão relativa aberrante por RT-qPCR dos novos transcritos intergênicos em amostras pancreáticas não-tumorais (NT), PDAC e PDX.	81
Figura 13 – Análise da sobrevida para TCONS00085964 e TCONS0003657482	
Figura 14 – validação da existência das novas formas de <i>splicing</i> (AS)	84
Figura 15 – A expressão relativa determinada por RT-qPCR dos lncRNAs confirma a expressão aberrante em tumor de PDAC dos lncRNA observada pelo RNAseq.	86
Figura 16 – <i>Heatmap</i> das expressões dos lncRNAs em linhagens celulares de PDAC.	88
Figura 17 - Análise de sobrevida para LINC01559, LINC02577 e UCA1.	89
Figura 18 – <i>Knockdown</i> dos lncRNAs em linhagem AsPC-1	90
Figura 19 – Ensaio de migração <i>transwell</i> e imageamento.	91
Figura 20 – Ensaio de invasão, <i>transwell</i>	92
Figura 21 – Avaliação da proliferação tumoral.	92

Figura 22 – Módulos de co-expressão e vias em que eles são enriquecidos. .	94
Figura 23 – <i>Heatmap</i> do módulo de coexpressão enriquecidos para a via de O-glicosilação de mucinas (<i>O-linked glycosylation of mucins</i>) com os lncRNAs DE do módulo.....	95
Figura 24 – <i>Heatmap</i> do módulo de coexpressão enriquecidos para a via de Reparo de DNA (<i>DNA repair</i>) com os lncRNAs DE do módulo.	96
Figura 25 – Expressão relativa de genes codificadores da rede de co-expressão do módulo de O-Glicosilação em <i>knockdown</i> de LINC01559 e CCAT1.....	99
Figura 26 – Expressão relativa de LINC01559, GALNT3, B3GNT3 e MUC13 em tumoresferas comparadas a células aderentes de AsPC-1 por RT-qPCR.....	101
Figura 27 – Expressão relativa dos genes codificadores do módulo de Reparo de DNA <i>knockdown</i> de UCA1.	102
Figura 28 – Ensaio cometa com <i>knockdown</i> de UCA1	103
Figura 29 – Crescimento tumoral para <i>knockdown</i> de UCA1.....	104

Lista de tabelas

Tabela 1 – Dados clínicos referentes as amostras de tumores.....	54
Tabela 2 – <i>Primers</i> usados na RT-qPCR dos lncRNA conhecidos DE	64
Tabela 3 – <i>Primers</i> usados na RT-qPCR dos novos transcritos DE.	64
Tabela 4 – <i>Primers</i> usados na RT-qPCR de genes codificadores avaliados. ..	65
Tabela 5 – <i>Primers</i> usados na PCR end point das novas formas de splicing DE	66
Tabela 6 – <i>Primers</i> usados na PCR end point dos novos transcritos intergênicos.	66
Tabela 7 – Número total de reads gerados pelo sequenciamento para cada biblioteca; número total e porcentagem <i>reads</i> mapeados na referência.	74
Tabela 8 – A reconstrução do transcrito pancreático.	75

SUMÁRIO

1. Introdução.....	23
1.1. Câncer.....	23
1.2. Câncer de Pâncreas.....	24
1.3. Lesões precursoras, diagnóstico e tratamentos.....	25
1.4. O transcriptoma eucariótico e estudos em câncer.....	28
1.5. Heterogeneidade tumoral e subtipos moleculares de PDAC.....	31
1.6. Microambiente tumoral (TME) de PDAC.....	37
1.7. Modelos experimentais de cultivo de PDAC.....	38
1.8. <i>Splicing</i> alternativo no câncer.....	40
1.9. RNAs longos não codificadores (lncRNAs).....	43
1.10. lncRNAs em PDAC.....	47
1.11. Redes de co-expressão gênica e WCGNA.....	51
2. Materiais e Métodos.....	53
2.1. Amostras de pacientes e xenotumores.....	53
2.2. Linhagens celulares, linhagem celular recém-derivadas de PDX e cultivo de tumoresferas.....	55
2.3. Extração de RNA.....	56
2.4. Preparo de bibliotecas e sequenciamento.....	57
2.5. Análises bioinformáticas do RNAseq.....	58
2.6. Análises por redes de co-expressão e WCGNA.....	61
2.7. Análises de sobrevivência.....	62
2.8. Síntese de DNA complementar por transcrição reversa (RT) seguida de reação em cadeia da polimerase quantitativa (qPCR).....	62
2.9. PCR <i>end point</i> seguida de eletroforese.....	66
2.10. Silenciamento transiente por siRNA.....	67

2.11.	Ensaio de proliferação e migração por imageamento ao vivo.	68
2.12.	Ensaio de migração e invasão <i>Transwell</i> [®]	69
2.13.	Ensaio xenográfico	70
2.14.	Ensaio cometa alcalino.....	71
3.	Objetivos.....	72
4.	Resultados.....	73
4.1.	A produção de bibliotecas de RNA-seq e sequenciamento preliminar 73	
4.2.	Sequenciamento HiSeq.....	74
4.3.	Análises do transcriptoma detectou novos transcritos, novas formas de splicing e lncRNAs diferencialmente expressos no PDAC	75
4.4.	Novos transcritos intergênicos detectados em PDAC	78
4.5.	Validação da expressão aberrante das novas formas de <i>splicing</i> (AS) em PDAC.....	82
4.6.	Validação da expressão aberrante de lncRNAs em PDAC	84
4.7.	lncRNAs envolvidos com migração de invasão	89
4.8.	Enriquecimento e WCGNA	93
4.9.	LINC01559 regulado pela expressão de CCAT1, regula genes da via de O-glicosilação e está enriquecido em CSCs.....	98
4.10.	UCA1 envolvido com reparo de DNA	101
4.11.	UCA1 diminui progressão tumoral em xenoinxertos	104
5.	Discussão	105
6.	Conclusão.....	114
7.	Referências.....	115
8.	Lista de anexos.....	144

1. Introdução

1.1. Câncer

Câncer é um termo genérico para um grande grupo de doenças que pode afetar qualquer parte do corpo, outros termos utilizados para nomeá-los são tumores malignos e neoplasias. A principal característica desse conjunto de doenças é a acelerada multiplicação de células atípicas desordenadamente, podendo invadir tecidos adjacentes do corpo e se espalhar para sítios distantes no corpo, processo denominado metástase, essa por sua vez é a principal causa de morte por câncer (Weinberg 2007).

O registro por escrito mais antigo do que hoje entendemos como câncer é uma descrição de um tumor de mama encontrado no Papiro Edwin Smith, escrito aproximadamente em 3000 AEC, nele o escritor descreveu como sendo um massa protuberante da mama, uma doença grave e sem tratamento (Hajdu 2011a; Breasted 1931). Já cerca de 1550 AEC, temos a primeira referência de um tumor de tecido mole, um lipoma, e outros registros que podem ser entendidos como hoje como possíveis cânceres de pele, útero, estômago e reto (Von Klein 1905). Galeno (130 a 200 DEC.), e outros médicos romanos, vai recomendar a remoção cirúrgica da massa tumoral como principal estratégia curativa, mas faz a ressalva de que em casos avançados deve-se evitar a remoção cirúrgica por dar como certa uma recidiva e invasão (Hajdu 2011a, 2011b). Apesar de existir um esforço curativo por estratégias medicamentosas, para alguns tipos de câncer como os cânceres pancreáticos, sobretudo o adenocarcinoma ductal (PDAC), a estratégia curativa principal ainda permanece sendo a ressecção cirúrgica.

O câncer é um dos maiores problemas de saúde pública em todo o mundo, inclusive sendo a segunda principal causa de morte no países desenvolvidos como Estados Unidos (Siegel, Miller, and Jemal 2016). A Organização Mundial de Saúde classifica o câncer como uma das *doenças não transmissíveis* (NCDs), as NCDs matam mais de 40 milhões de pessoas a cada ano. Desse total de vítimas 15 milhões têm entre 30 e 69 anos (WHO 2018), os chamados de casos “prematureos”, que são prevalentes em países de baixos

rendimentos (PBR) como os países de rendimento médio (PRM) (F. Bray et al. 2018), como o Brasil.

1.2. Câncer de Pâncreas

O câncer de pâncreas (PC) é a mais mortal malignidade, e certamente uma das doenças mais devastadoras uma vez que apresenta o pior resultado de sobrevivência quando comparada a qualquer tipo de câncer humano c, sendo apontado como o décimo tipo de câncer mais comum, quarta *causa mortis* por câncer nos Estados Unidos com projeções de vir a ser a segunda maior causa de morte por câncer até 2030 e terceira causa de morte relacionada ao câncer na União Europeia até 2025 (Lola Rahib et al. 2014; RL Siegel 2017). No Brasil, embora não seja uma das neoplasias mais prevalentes, figura como sendo responsável por 2% de todos os casos de câncer e 4% do total de mortes por essa doença, ele figura como a oitava causa de morte por câncer, dado o diagnóstico geralmente em fase localmente avançada ou metastática da doença (Soldan 2017). As taxas de incidência são diferentes para dois grupos etários: de 10/100.000 habitantes entre pessoas de 40 a 50 anos de idade e de 116/100.000 para pessoas entre 80 e 85 anos, além da idade tem-se que existe uma incidência é mais significativa no sexo masculino (INCA 2018).

Correlacionam-se com a incidência do PC alguns atores de risco, tais como hábito de tabagismo, sedentarismo, dieta rica em gordura e colesterol; as ocorrências de patologias como pancreatite crônica, a cirrose, a obesidade, diabetes mellitus; questões ambientais como a exposição ocupacional aos agentes carcinógenos; baixo nível socioeconômico. Além dessas algumas condições hereditárias tem correlação com PC tais como ascendência judaica (Ashkenazi); síndromes familiares de pancreatite hereditária, câncer colorretal não polipoide hereditário, câncer de mama e ovário hereditários, melanoma múltiplo atípico familiar, Peutz-Jeghers e ataxia-telangectasia (Kushi et al. 2012).

O PC constitui um conjunto heterogêneo de tumores caracterizados pela alta agressividade e mortalidade. Dentre esses tumores os adenocarcinomas ductais (PDAC) que acometem a região exócrina do órgão constituem o tipo mais frequente de neoplasia pancreática, representando 85% dos casos (DP Ryan 2014). E estudos estimam que o percurso do tumor de pâncreas é desenvolver-

se por 10 a 15 anos praticamente assintomático até que esteja avançado e tenha metastizado (Corbo, Tortora, and Scarpa 2012). No quadro geral observa-se uma sobrevida mediana de menos de 12 meses após o diagnóstico, entretanto existe uma taxa de 10% cuja sobrevida observada foi de 5 anos, daí o PDAC apresentar-se como um grande desafio para a saúde ocidental (Vincent et al. 2011; RL Siegel 2017). Um caso muito representativo sobre a falta de tecnologias e diagnóstico foi descrito em 2014 por Siriwardena:

Um homem de 74 anos vem se automedicando com antiácidos por causa de recorrentes indigestões, dado a persistências dos sintomas por três meses ele consultou um Clínico Geral e este o receitou um inibidor da bomba de prótons e marcou o retorno em quatro meses. Dois meses depois ele desenvolveu dor nas costas, e lhe foi prescrito paracetamol e uma movimentação correta, passadas quatro semanas ele desenvolveu icterícia e urina escura, uma tomografia de contraste computadorizada confirmaria um tumor avançado e não removível câncer de pâncreas. Foi-lhe prescrita quimioterapia paliativa, mas ele veio a morrer em 5 meses (Siriwardena and Siriwardena 2014).

Devido à supracitada falta de sintomas precoces, os pacientes geralmente são diagnosticados com a doença localmente avançada, ou metastática, quando a estratégia de remoção cirúrgica do tumor é inviável, demonstrando a necessidade de tecnologias de diagnóstico, prognóstico e tratamento do PC.

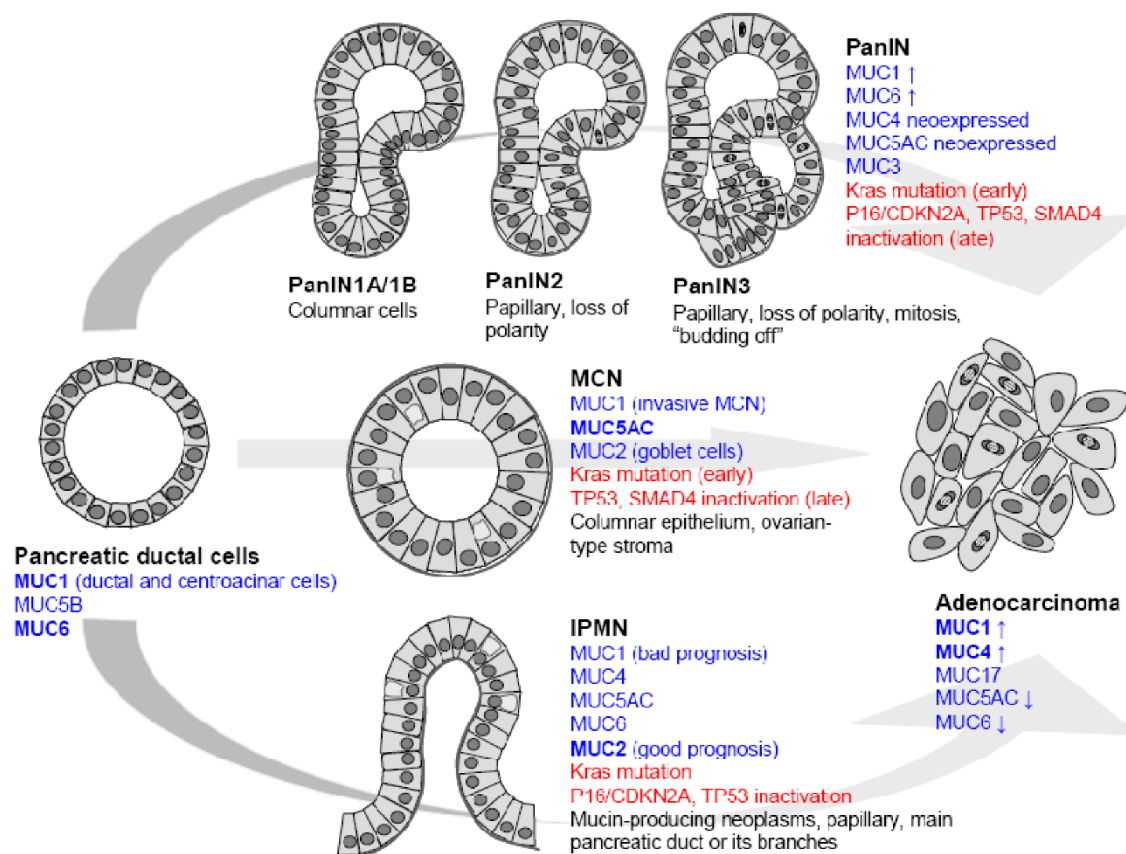
1.3. Lesões precursoras, diagnóstico e tratamentos.

Indícios fortes levam a crer que a maioria dos casos de PDAC originam-se a partir de lesões microscópicas (Figura 1) precursoras do câncer *in situ*. Tal lesão é denominada de neoplasia intraepitelial pancreática (PanIN). A maior dificuldade para um diagnóstico precoce desse tipo de lesão deve-se ao fato de estarem abaixo do tamanho detectável pelas técnicas clínicas de imagem, como a tomografia computadorizada (CT), ultrassonografia (US) e a ressonância magnética (RM) abdominal pois essas técnicas são ineficazes para a detecção lesões menores que 1-2 cm (em seus maiores eixos). Isso torna o diagnóstico e tratamento de lesões em pâncreas dificultosos, apesar de ocasionalmente acontecer a detecção de PanIN em cerca de 3% dos indivíduos, são detecções acidentais. Além das PanINs existem outros tipos de lesões pancreáticas e elas

também são considerados precursoras do PC, são as neoplasias mucinosas papilares intraductais (IPMNs) e as neoplasias císticas mucinosas (MCN). Essas lesões possuem um potencial de malignização variável, sendo que o grau de displasia celular e a localização anatômica da lesão no pâncreas são considerados fatores preditores de malignização. Os PDAC que surgem dessas lesões precursoras são geralmente indistinguíveis histologicamente, mesmo se originados de tipos diferentes de lesões, embora o tipo histológico chamado de não-cístico mucóide ou colóide seja descrito como um estado cístico precursor. Quanto a IPMNs são frequentemente encontradas associadas ao PDAC avançado, o que levanta duas hipóteses sobre o IPMN, uma propõe que o IPMN origina o PDAC e outra que PDAC e IPMN coexistem e crescem em paralelo no pâncreas (Basturk et al. 2015; Hruban 2004; WR Brugge 2004; Brand et al. 2007; Balachandran 2017; A Scarpa 2018).

O diagnóstico inicial do PDAC é feito através dos sinais clínicos e radiológicos e confirmado por análise histopatológica do tecido coletado por biópsia por agulha fina. O marcador sérico CEA e o CA 19-9 são comumente utilizados no processo de diagnóstico, prognóstico e monitoramento pós-cirúrgico em contexto de PDAC (Poruk et al. 2013; Santucci et al. 2018; Jiarui Li et al. 2019; Locker et al. 2006). Entretanto esses marcadores tem limitações sérias, o CA 19-9 é ausente em pessoas Lewis negativas (Berger et al. 2008), uma outra limitação é o aumento dos níveis séricos em casos de pancreatite crônica e outros tipos de câncer do sistema digestivo (Slesak et al. 2000; Mann et al. 2000; Biswas et al. 2019). A grande maioria dos casos de PDAC são diagnosticados quando incuráveis e inoperáveis, com extensa invasão local ou metástase de órgãos distantes, apenas 15% dos casos diagnosticados podem ser removidos cirurgicamente (Stathis and Moore 2010). Estima-se que a detecção precoce seja capaz de aumentar a sobrevivência dos pacientes em 30 a 40% (L Rahib 2014; Seufferlein and Mayerle 2016).

Figura 1 – Evolução das lesões precursoras para o PDAC.



Modelo de malignização das lesões precursoras PanIN, MCN e IPMN para o adenocarcinoma ductal de pâncreas (PDAC). Para cada tipo de lesão estão destacadas as (fonte azul) as mucinas avaliadas como marcadores por imunohistoquímica e os genes comumente mutados (fonte vermelha). Fonte: (Jonckheere, Skrypek, and van Seuningem 2010)

Nos casos onde é possível a ressecção cirúrgica do PDAC, o tratamento padrão, "tamanho único", é tratamento com análogo de nucleosídeo gemcitabina como tratamento adjuvante padrão (ESPAC-3) (Neoptolemos, Stocken, 2010). Mais recentemente regimes quimioterápicos de nab-paclitaxel (inibidor de despolimerização de microtúbulos – paclitaxel associado a albumina humana) e FOLFIRINOX (ácido folínico, fluorouracila, inibidor de topoisomerase – irinotecano, oxaliplatina) vêm sendo administrados mas são limitados, acompanhado por efeitos colaterais tais como náuseas, neutropenia, neuropatia e complicações infecciosas (Singh, Hasselluhn, and Neesse 2019; Neoptolemos et al. 2018).

São múltiplos os fatores descritos contribuem para severidade dos prognósticos e do elevado índice de mortalidade, quando comparados ao de incidência (praticamente 100%), mas podem ser agrupados em dois problemas:

diagnóstico tardio e resistência ao tratamento, até o presente momento carente de tecnologias efetivas de estratificação para tratamento personalizado (Seufferlein and Mayerle 2016).

A heterogeneidade do PC torna as opções terapêuticas baseadas na abordagem de “tamanho único” ineficaz. O diagnóstico e estratificação de malignidades é uma estratégia de sucesso em outros tipos de câncer e poderiam beneficiar de tratamentos mais específicos reduzindo as abordagens por terapias fúteis para PDAC, conseqüentemente observaríamos uma melhoria dos desfechos e da qualidade de vida daqueles cuja sobrevida, até agora, a longo prazo, não é promissora (Torres and Grippo 2018). Um exemplo promissor da estratégia de identificação de perfis moleculares pode ser observado no estudo que tratou casos metastáticos de PDAC com mutação nos genes BRCA1 ou BRCA2, a ocorrência dessa mutação germinativa em pacientes com PDAC é da ordem de 4 a 7% (Golan et al. 2018; Grant et al. 2015; Friedenson 2005), o uso do fármaco olaparibe (inibidor da enzima poli ADP-ribose polimerase-1 – PARP1) resultou num aumento da sobrevida e menor progressão tumoral (Golan et al. 2019). Esse medicamento no Brasil recebe o nome de Lynparza™ Comprimido, na Anvisa já é registrado para o tratamento de câncer de mama (Anvisa RE 100/2017) e foi ampliado para o uso de câncer de ovário (Anvisa - RE 333/2019), porém até o dado momento não foi encontrado nenhuma solicitação de registro ou ampliação do uso para PDAC no sistema da Anvisa.

1.4. O transcriptoma eucariótico e estudos em câncer

Com a conclusão do Projeto Genoma Humano (HGP) foram anotadas as sequências de DNA armazenados nos cromossomos, a isso denomina-se genoma humano. Tal fato possibilitou explorar a organização e composição genética humana com um grande avanço da biologia molecular (Lander et al. 2001; F. S. Collins, Morgan, and Patrinos 2003). Graças ao avanço paralelo das tecnologias de sequenciamento e bioinformática este campo se consolidou como sendo crucial para: uma maior compreensão da biologia molecular, para o desenvolvimento e manipulação de tecnologias genéticas, o estabelecimento da medicina personalizada para tratar no tratamento do câncer (Mukherjee 2019; Friedman et al. 2015).

O RNA tem um papel chave e intermediário entre o genoma e o proteoma (Conesa et al. 2016). O conjunto de todos RNAs de uma célula constitui o transcriptoma, que pode ser definido como a soma de toda a porção genética correspondente que atua como molde para síntese de qualquer RNA pela célula, numa dada quantidade, em uma etapa específica do desenvolvimento ou condição fisiológica do indivíduo (Z. Wang, Gerstein, and Snyder 2009). O transcriptoma é um estado específico da expressão do repertório gênico de uma célula, logo conhecê-lo é interpretar o funcionamento do genoma, o controle operacional molecular das células e tecidos, isso faz do transcriptoma objeto de estudo eficaz para entender o desenvolvimento de algumas doenças celulares como câncer (Z. Wang, Gerstein, and Snyder 2009; Wickramasinghe et al. 2014).

Os RNAs são classificados em classes funcionais e estruturais. Algumas classes de RNAs são envolvidos na tradução: RNAs transportadores (tRNA), RNAs ribossômicos (rRNA), pequenos RNAs nucleares envolvidos em eventos de *splicing* (snRNP), e pequenos RNAs nucleolares envolvidos principalmente em a modificação de outros pequenos RNAs (snoRNA) como RNAs ribossômicos e RNAs transportadores. Outras classes de RNAs são envolvidos em processos regulatórios: RNAs reguladores curtos, pequenos RNAs endógenos de interferência (endo-siRNA) e microRNAs (miRNA), uma outra classe é a dos RNAs não codificadores cujo comprimento é superior a 200pb, chamados RNAs não codificadores longos (lncRNAs), o quais são classificados prioritariamente pela seu tamanho, sabe-se que são envolvidos em diversos papéis biológicos (Hombach and Kretz 2016).

O catálogo de referência GENCODE (GRCh38.p12) descreve o transcriptoma humano como sendo um conjunto de 226882 transcritos diferentes, produtos de um total de 60603 genes diferentes. No conjunto de transcritos as duas classes majoritárias de RNAs são os lncRNAs e mRNAs, apesar dos mRNAs serem a classe mais estudada ambas as classes de RNA tem representatividades semelhantes no transcriptoma, apesar disso o conhecimento sobre os lncRNAs está na infância quando comparado aos mRNAs. O catálogo do GENCODE descreve que 75057 (33%) dos transcritos, produtos de 16840 (27%) genes, são anotados como sendo lncRNAs; frente aos

83660 (36%) transcritos, produto de 19975 (32%) genes, são anotados como codificadores de proteínas (Frankish et al. 2019).

São várias as tecnologias de estudo e análise do transcriptoma, todas elas objetivando determinar prioritariamente a sequência primária e a quantificação de vários transcritos ou genes em uma mesma análise. As tecnologias bioquímicas empregadas nessas técnicas são baseadas na adição de nucleotídeos modificados, hibridização e sequenciamento por síntese. O método mais comum baseado em sequências específica é microarranjo (*microarray*), surgido em 1999 e utilizado até hoje, nele sondas complementares para determinadas regiões alvo produzindo resultados quantitativos, este método possui algumas limitações, como o conhecimento prévio da sequência do genoma a ser analisada, altos níveis de ruídos de fundo devido a hibridização cruzada, além de ser difícil comparar os níveis de expressão entre diferentes experimentos e exigir métodos de normalização dos dados específicos. Outras técnicas são baseadas no sequenciamento Sanger; como CAGE (*Cap Analysis of Gene Expression*) e SAGE (*Serial Analysis of Gene Expression*), ambas são exclusivamente para o sequenciamento e dependem da construção de bibliotecas de concatâmeros com insertos de cDNA e amplificação em bactérias para posterior sequenciamento Sanger (Shiraki et al. 2003; Kodzius et al. 2006; Shafee, Lowe, and Lowe 2017; Yuan et al. 2001). Enquanto CAGE e SAGE tem poder exploratório para estudar um conjunto de sequências de regiões desconhecidas, o microarranjo de DNA permite estudar a expressão de genes, mapear no genoma os locais de ligação de proteínas que interagem com o DNA e examinar outras interações de longo alcance do DNA.

A maior robustez na para identificar o conjunto completo de um transcriptoma, RNAs longos, curtos, novos transcritos (genes não anotados), novas formas de splicing, transcrições de fusão gênica, sobreveio em 2004 com o desenvolvimento de tecnologias de sequenciamento de nova geração (NGS), ou sequenciamento de RNA (RNA-seq), capaz de revelar a complexa e dinâmica realidade do transcriptoma com o maior nível de sensibilidade e precisão. Atualmente a metodologia de semicondutores (Ion), Sequenciamento de moléculas únicas (Pacific Biosciences e o Oxford Nanopore) e principalmente Sequenciamento por síntese, da plataforma illumina são predominantes na literatura, destacando-se que a plataforma Illumina é a mais bem estabelecida.

Porém alguns desafios devem ser considerados, como a necessidade de computadores com grande capacidade de armazenamento de dados, desenvolvimento de técnicas de normalização de dados e métodos estatísticos específicos para essa análise (Wickramasinghe et al. 2014; W. Zhao et al. 2014).

Uma vez que o câncer é uma doença fundamentalmente genômica, e conseqüentemente transcriptômica, o RNA-seq tão logo tornou-se crucial para o estudo e entendimento dos caminhos moleculares do câncer (Weinstein et al. 2013; Marguerat and Bähler 2010; Maher et al. 2009), inclusive sendo um divisor de águas para as pesquisas sobre o câncer de pâncreas (Müller et al. 2015). O uso de NGS além da pesquisa tem sido utilizado de forma translacional na clínica para guiar decisões no tratamento de pacientes com cânceres mais bem estabelecidos como câncer de mama (Freedman et al. 2018).

Os dados da literatura de RNA-seq de cânceres possuem um viés importante, são predominantemente oriundos do estudo da fração poliadenilada do transcriptoma (poliA+) em detrimento da fração não-poliadenilada do transcriptoma (poliA-). Pois, tem-se duas abordagens técnicas no preparo das bibliotecas para o sequenciamento, por possuir uma etapa seleção positiva de RNA poliA+, partindo-se o RNA total, acaba enriquecida para esses nas bibliotecas geradas, a outra é chamada RNA total depletada de RNA ribossomal por possuir uma seleção negativa, na qual, partindo-se o RNA total, existe uma remoção dos rRNA, os quais são altamente abundantes, compondo assim uma amostra final contendo a fração poliA+ e poliA- do transcriptoma. O sequenciamento de bibliotecas oriundas da estratégia de depleção do rRNA demonstra ser mais enriquecida para lncRNAs e pequenos RNAs quando comparada a outra estratégia, apesar daquela apresentar maior concentração de seqüências intrônicas de mRNAs imaturos (S. Zhao et al. 2018).

1.5. Heterogeneidade tumoral e subtipos moleculares de PDAC

A heterogeneidade molecular do câncer fica evidente na presença de translocações cromossômicas, deleções de genes supressores de tumor e anormalidades cromossômicas distintas entre tumores histologicamente semelhantes. A classificação em subtipos do câncer humano vem sendo cada

vez mais informativa e clinicamente útil à medida que as análises moleculares mais detalhadas dos tumores têm sido realizadas (Alizadeh et al. 2000).

Como já dito no item 1.1, encontramos na literatura casos de tumores sólidos histopatologicamente idênticos, isto é, tumores diagnosticados como sendo a mesma malignidade que demonstram uma significativa heterogeneidade na resposta aos tratamentos. Tal heterogeneidade pode ser determinada pelas distinções dos mecanismos moleculares intrínsecos entre os casos da mesma malignidade, revelando a importância de outras características, que não apenas histológicas e citológicas, como características moleculares para distinguir entidades de cânceres até então microscopicamente indistinguíveis. Ao fazê-las distintas torna-se possível distinguir em subclasses, ou subtipos tumorais, o que permite averiguar os melhores resultados de tratamento através do direcionamento de terapias e medicamentos para tais subtipos de tumores nos quais eles são seletivamente eficazes (Collisson et al. 2019). Um exemplo de sucesso dessa abordagem está nos tumores de mama (subtipo HER2 amplificado), de pulmão (subtipo câncer de células não-pequenas com mutações específicas no gene EGFR), o primeiro responde melhor ao tratamento com Tamoxifen e o segundo tem melhor resposta a inibidores de tirosina quinase, gefitinibe (Slamon et al. 2001; Lynch et al. 2004).

Histologicamente os casos de PDAC são caracterizados por serem compostos de células bem diferenciadas de adenocarcinoma ductal e, eventualmente a presença de células glandulares mergulhadas em um microambiente tumoral (TME) abundante constituído de compostos acelulares como colágeno, ácido hialurônico, fibronectina, matriz extracelular. Outros tecidos como de células inflamatórias, macrófagos associados a tumores, neutrófilos associados a tumores, Células T, fibroblastos associados ao câncer e células neurais se acumulam na neoplasia (Neesse et al. 2019).

Há um esgotamento das tecnologias baseadas nas características histopatológicas do PDAC não se mostraram úteis para escolha de estratégias terapêuticas (Singh, Hasselluhn, and Neesse 2019). Por outro lado, a análise global da expressão gênica, o transcriptoma, combinada com dados clínicos e patológicos, tem se mostrado útil para a identificação de subtipos de muitos tumores humanos (Alizadeh et al. 2000). Com o sequenciamento do genoma humano e as ferramentas de NGS foi possível sub-classificar os tumores com

base em mutações nas principais vias de sinalização, incluindo KRAS, reparo do dano ao DNA, remodelamento da cromatina e outros (Bailey et al. 2016; Notta et al. 2016; Witkiewicz et al. 2015; Waddell et al. 2015; Biankin et al. 2012; Connor and Gallinger 2017). Incapazes de caracterizar subtipos moleculares mas serem um dos componentes para tal classificação as mutações somáticas são recorrentemente descritas, tais como KRAS, TP53, CDKN2A, SMAD4, RNF43, ARID1A, TGF β R2, GNAS, RREB1 e PBRM1 (B. J. B. Raphael et al. 2017; Bailey et al. 2016; Waddell et al. 2015; Witkiewicz et al. 2015).

Vários estudos examinaram a expressão gênica usando microarranjo, sequenciamento de RNA identificaram distintas assinaturas de RNA de PDAC. Em 2011, Collisson *et al* através da análise por NMF de dados de expressão gênica de mRNA de 85 tumores humanos não tratados, associada aos dados clínicos correspondentes propôs a existência de três subtipos de PDAC: clássico, quase mesenquimal e *exocrine-like*. Respectivamente o subtipo clássico apresenta alta expressão de genes epiteliais e genes associados à adesão, o subtipo quasimesenquimatoso tem alta expressão de genes associados ao mesênquima, enquanto o subtipo *exocrine-like* apresentou elevada expressão de genes de enzimas digestivas (Collisson et al. 2011).

Utilizando um número maior de amostras primárias, 145 tumores cirurgicamente removidos, e com um conjunto de amostras metastáticas, 61 amostras, 17 linhagens celulares, 46 amostras de pâncreas e 88 amostras normais adjacentes de sítios distantes, Moffitt *et al* propôs por NMF a estratificação dos componentes tumorais e componentes do estroma. Foi proposto dois subtipos tumorais; clássico e *basal-like*, e dois subtipos de estroma: normal e ativado. Segundo os resultados da pesquisa as três classificações propostas por Collisson *et al*. Anteriores apesar de serem observadas nos dados do grupo não tinham poder prognóstico. Os genes da assinatura tida como clássico de Collisson sobrepõem-se aos do subtipo clássico de Moffitt, motivo pelo qual o grupo manteve a nomenclatura de “clássico” para esse subtipo. Quanto a assinatura gênica associada ao subtipo quasimesenquimatoso era uma mistura entre subtipo tumorais e subtipos estromais. Tumores tidos como basais e clássicos foram identificados pelo modelos propostos nos subtipos de estroma normal e estroma ativado (Moffitt et al. 2015).

Em 2016, Bailey *et al*, investigando os perfis de expressão de mRNA por microarranjo de 266 tumores primários não tratados e RNAseq de 96 amostras, as quais continham espécies histopatologicamente subclassificadas como coloidais, adenoescamosas, PDAC associadas a IPMN, tumor acinar e algumas outras variantes raras de PC. Este trabalho propôs a estratificação em quatro subtipos moleculares: escamoso, progenitor pancreático, imunogênico e exócrino-endócrino anormalmente diferenciado, semelhante ao *exocrine-like*, ADEX (Bailey et al. 2016).

Posteriormente um estudo colaborativo publicado pelo TCGA em 2017 teve resultados semelhantes aos resultados de de Moffitt *et al*. Utilizando-se RNAseq, de bibliotecas de seleção de transcritos poli(A), integradas a análises genômicas histopatológicas e proteômicas, o trabalho propôs para um n = 150 amostras tumorais primárias uma estratificação em dois subtipos baseados nos perfis de expressão gênica são eles *basal-like* (ou escamoso) e clássico (ou progenitor). Os tumores clássicos demonstraram ser enriquecidos para mutações em GNAS enquanto nos tumores *basal-like* existe a prevalência de mutações em TP53. O trabalho aponta que o perfil ADEX, proposto por Bailey *et al*., trata-se de um viés de amostras de baixa celularidade neoplásica, por isso não se trata de um subtipo de PDAC. Analisando-se a coorte para a assinatura dos 360 lncRNAs mais DE, foi confirmada que os subgrupos têm uma perfeita sobreposição com subtipos clássico, cujos lncRNAs EVADR, DEANR1 e GATA6-AS1 foram tidos como uma assinatura de DE associada, e o subtipo *basal-like*. Este estudo revelou pela primeira vez a existência de lncRNAs nas assinaturas de expressão gênica que definem subtipos moleculares, indicando a contribuição dessa classe de transcritos para processos malignos. Foi notado também que a classificação das amostras a partir da expressão de lncRNAs se mostrou independente da celularidade tumoral, apontando o valor dessa classe de transcritos para a classificação molecular de tumores (B. J. Raphael et al. 2017).

Em 2018, Puleo *et al* partindo de 305 amostras tumorais fixadas em formalina, sem a presença de amostras não tumorais. Foram realizados sequenciamento de painéis de DNA, avaliação transcriptômica por microarranjo e análises imuno-histoquímicas. Foram feitas análises de componentes independentes para deconvolução para o estabelecimento e identificação de

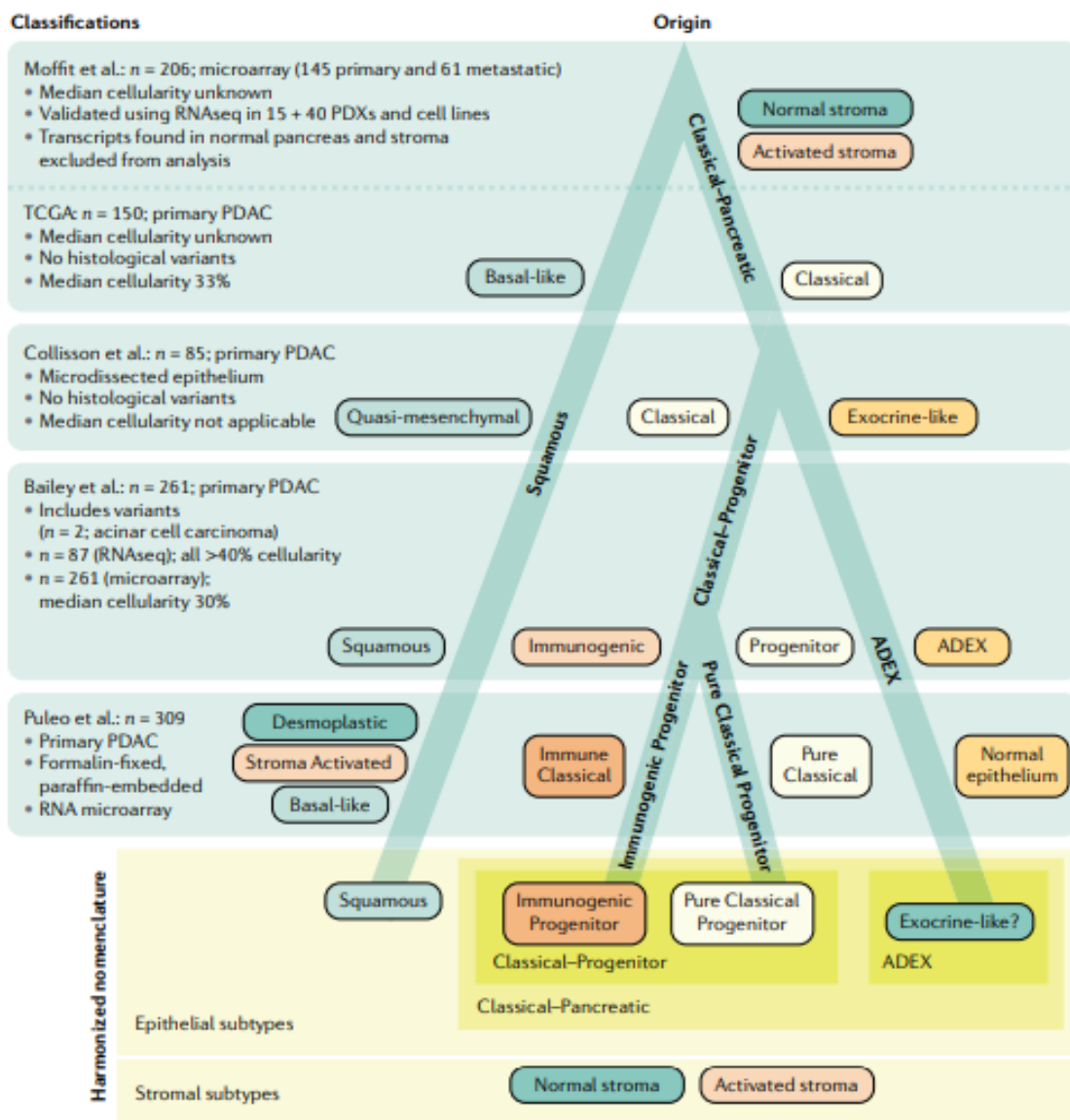
assinaturas normais, tumorais e microambientes das amostras. O estudo conclui que o proposto subtipo ADEX trata-se de um viés amostral, neste caso a hipótese é ser o resultado de amostras de PDAC com contaminação de células acinares pancreáticas, por outro lado foram encontradas grandes semelhanças com as classificações já propostas para amostras de PDAC com alta celularidade tumoral, *basal-like* e clássico, porém somadas as análises imunohistoquímicas foi proposta a estratificação em 5 subtipos, chamados de basal puro, estroma ativado, desmoplásico, clássico-puro e clássico-imunológico, todos eles com potencial prognóstico. Os subtipos clássico-imunológico e desmoplásico são enriquecidos para CTLA4 tornando-os potencialmente mais apropriados para a terapia anti-CTLA4 com ipilimumabe (Puleo et al. 2018). O ipilimumabe por é um fármaco de imunoterapia já utilizado para o tratamento de melanoma (Hodi et al. 2010). Os subtipos desmoplásicos, basal puro, estroma ativado são tem alta expressão dos genes CD276 e HAVCR2 com alto potencial terapêutico. Já todos os subtipos excetuando-se o clássico-puro possuem um padrão específico de expressão de PDL2 que, em geral, não regulam positivamente qualquer um dos pontos de verificação imunológicos (Puleo et al. 2018).

Wartenberg *et al.*, em 2018, fez uma proposta translacional de classificação de 110 amostras de PDAC integrando características moleculares e biológicas do tumor e do TME. Foram descritos três subtipos: escape imunológico, imune-enriquecido, imune-exaurido. Respectivamente as assinaturas de cada subtipo são caracterizadas por serem pobres em células B e T, enriquecidos em células T_{reg} FOXP3⁺; baixa frequência de mutações em CDKN2A, SMAD4 e PIK3CA; esse subtipo demonstrou prognóstico ruim. O outro subtipo, o imune-enriquecido, apresenta uma maior número de células B e T, menor número de células T_{reg} FOXP3⁺; com baixo grau de brotamento; baixa frequência de mutação em CDKN2A e PIK3CA, porém uma frequência maior de mutações em genes de resposta ao dano ao DNA (STK11, ATM); presença de subpopulação de tecido linfóide terciário; apresenta um prognóstico melhor que o subtipo anterior. Já o terceiro subtipo, imune-exaurido, é caracterizado por TME imunogênico, nele encontramos dois grupos: um com alta expressão de PD-L1 e alta taxa de mutação PIK3CA, e o outro grupo com

microssatélite-instável e alta prevalência de mutações em JAK3 (Wartenberg et al. 2018).

Em 2019 Collisson *et al*, elaborou uma proposta que condensa comparada e organiza a estratificação em subtipos de PDAC mais relevantes em uma árvore filo-transcriptômica (Figura 2).

Figura 2 – Árvore filo-transcriptômica de subtipos de expressão molecular de PDAC.



Fonte: (Collisson et al. 2019)

Recentemente Maurer *et al*, utilizando 60 amostras de PDAC separando por microdissecção a laser o estroma adjacente do epitélio neoplásico associado realizaram RNAseq pareado para gerar expressão gênica pareada e analisando

pelo algoritmo ADVOCATE de aprendizado de máquina obteve como resultados a estratificação em subtipos semelhantes aos já descritos, propondo assim como Collisson um sistema capaz de harmonizar as outras propostas as estratificações do PDAC. O resultados desse estudo concluíram dois subtipos de epitélio neoplásico, denominados de subtipos epiteliais clássico e basal, mantendo a nomenclatura Moffit. Por outro lado, as análises de NMF apesar de gerarem também dois grupos com assinaturas estromais distintas, os resultados levaram a uma nova estratificação de subtipos em subtipo *imune-enriquecido*, caracterizado por serem enriquecidos para vias de sinalizações imunes e interleucinas; e o subtipo *enriquecido em ECM*, caracterizado pelo enriquecimento de vias associadas à matriz extracelular. Por fim foi demonstrado que a combinação desses subtipos tem significativa alteração para sobrevida analisados em bancos de dados do ICGC, TCGA e UNC (Maurer et al. 2019).

1.6. Microambiente tumoral (TME) de PDAC

O TME do PDAC é caracterizado por ter uma proeminente reação desmoplásica, com predominância de células de um estroma fibrótico denso perante 5% a 20% de células neoplásicas (Iacobuzio-Donahue et al. 2002; Wood and Hruban 2012). O PDAC é marcado por uma heterogeneidade e um intrincado ecossistema de alta complexidade entre os diferentes componentes e tipos celulares, incluindo as populações de células estreladas pancreáticas (PSCs), fibroblastos associados ao câncer (CAFs), células imunes, vasos sanguíneos tumorais e nervos, componentes não celulares, como colágenos, fibronectina, glicoproteínas, proteoglicanos, ácido hialurônico, citocinas e fatores de crescimento (Neesse et al. 2019; Mekapogu et al. 2019; Feig et al. 2012). O comportamento do TME é plástico, dinâmico e atuante em promover a progressão do tumor, através da formação de nichos que modulam o ambiente imune supressivo, crescimento tumoral, comportamento invasivo, a metástase e apontado como envolvido na resistência terapêutica. (Seo and Pillarisetty 2017; van Dalen et al. 2018). Em geral, os tumores pancreáticos são pouco vascularizados e ECM densa, a alta pressão do tecido displásico acaba por induzir a vasoconstrição, que por sua vez diminuiu vertiginosamente a perfusão do tumor e resposta terapêutica. Alguns estudos mostraram que mirar as vias

epigenéticas poderia ajudar na reprogramação do microambiente tumoral em detrimento do tumor com resultados positivos para a diminuição da progressão tumoral (Mekapogu et al. 2019; Nguyen et al. 2017; Yamamoto et al. 2016).

Apesar das heterogeneidades relatadas por diversas abordagens, os estudos pela abordagem de NGS do transcriptoma do PDAC e consequentemente do TME, estroma, tem sido eficaz em agrupar em dois subtipos normal e ativado (Maurer et al. 2019; Collisson et al. 2019).

1.7. Modelos experimentais de cultivo de PDAC

Um dos gargalos nas pesquisas com tumores, inclusive o PDAC cujo estroma é muito abundante, está nas limitações intrínsecas de cada modelo experimental do cultivo de câncer o que dificulta a dissecação de vias e dados moleculares, por exemplo isolar mecanismos observados por em estudos exploratórios obtidos de amostras cirúrgicas. As culturas convencionais de células em 2D, monocamada, apesar de conservarem o fenótipo tumoral, não são capazes de mimetizar a complexidade e heterogeneidade dos tumores, uma vez que *in vivo* eles crescem em uma conformação tridimensional (3D) com uma organização e arquitetura específicas, o que, obviamente, uma cultura de células 2D não é capaz de reproduzir. Consequentemente, a incapacidade de recapitular a complexidade tumoral e estromal faz com que numerosos sinais que governam diferentes processos celulares sejam perdidos quando as células são cultivadas em substratos plásticos 2D (Zanoni et al. 2016; Lee et al. 2007; Nyga, Cheema, and Loizidou 2011; Cukierman et al. 2001). Outros modelos de cultivo tumoral *in vitro* e *in vivo* possuem condições ambientais para que as células em cultivo construam algum grau de maior organização da complexidade tumoral em 3D, podendo inclusive ocorrer a organização de algum estroma a depender do modelo, tudo isso para que haja uma maior recapitulação possível do contexto do tumor *in situ*.

In vitro, o modelo de co-cultivo, 2D, no qual existem o cultivo de linhagens tumorais juntamente a outras linhagens celulares, como de fibroblastos. Esse sistema mimetiza parte da sinalização entre tumor e CAF (estroma) sendo utilizado para estudar a interação tumor-estroma, as maiores limitações desse modelo são a pouca heterogeneidade, falta de microambiente

tumoral, uma seleção agressiva de clones, desvio genético(Q. Sun et al. 2018). O grupo de *Wei et al.* utilizando esse sistema determinou especificamente elementos do *crosstalk* tumor-estroma da via SDF-1/SATB-1, mostrando que essa está diretamente envolvida na progressão tumoral bem como resistência a gemcitabina do PDAC (Wei et al. 2018).

Por outro lado, o cultivo tumoral *in vitro* existem abordagens 3D por modelos: esferoide multicelulares de tumor obtido por cultura de linhagens celulares em condições não aderentes; as tumor-esferas, mais utilizadas como um modelo de expansão de células-tronco tumorais (CITs) através do cultivo em um meio de cultivo sem soro e suplementado com fatores de crescimento; esferas tumorais derivadas de tecido e esferóides multicelulares organotípicos, obtidos por dissociação mecânica do tecido tumoral (Weiswald, Bellet, and Dangles-Marie 2015). O modelo de organoide, apesar de ter surgido para o estudo de tecidos saudáveis, também é muito utilizado para o estudo de PDAC, esse sistema é baseado em cultivo de uma fração tumoral em meio de cultivo que permita uma certa arquitetura semelhante ao órgão de origem. Podem ser construídas a partir de biópsias, constituindo o organoide derivado de paciente (PDO), tem grande poder para o estudo do *crosstalk* entre PDAC e CAF, bem como outras células, características moleculares do início da progressão tumoral, invasão e migração. A literatura aponta como limitações do PDO a dificuldade de estabelecer replicatas pelo número insuficiente de células estreladas primárias provenientes da biópsia, a falta de modelos para estágios avançados do organoide além da grande dificuldade técnica (Q. Sun et al. 2018). A identificação do mecanismo de reorganização da ECM para a invasão de PDAC por Endo180 foi identificado em modelos PDO, basicamente Endo180 está relacionado com a capacidade invasiva das PSCs que abrem caminho para as células de tumorais através da fosforilação da cadeia leve da miosina 2 (MLC2) (Koikawa et al. 2018).

Entretanto a maior recapitulação da complexidade tumoral é observada nos modelos *in vivo*, os quais têm a vantagem de fornecer um microambiente 3D próprio e suplementar nativo do próprio hospedeiro. O modelo *in vivo* mais utilizado rotineiramente é o xenoenxerto de tumores humanos (PDX) ou linhagens tumorais (Xenoenxertos de co-injeção derivados de linhagem celular) e os modelos de camundongos geneticamente “engenheirados”. Ambos, PDX e

xenoinxertos, consistem em inocular pequenos fragmentos de tumor ou um conjunto de células tumorais em um modelo animal, como por exemplo o camundongo (*Mus musculus*) e permitir o desenvolvimento ectópico tumoral. Uma grande vantagem desse método é a capacidade de transplantar tanto o tecido do câncer quanto o estroma circundante para imitar a complexidade do ambiente do tumor humano (Zanoni et al. 2016), por outro lado apresenta as limitações de ter uma seleção clonal intrínseca de populações tumorais mais agressivas, a seleção de pacientes (no caso do PDX) e baixa taxa de viabilidade dos enxertos (Q. Sun et al. 2018).

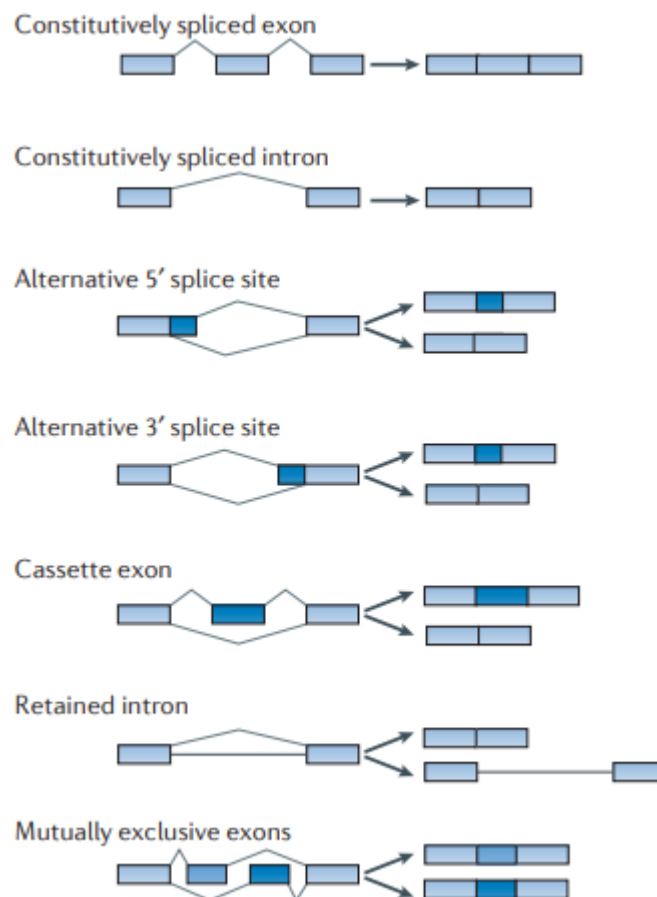
1.8. *Splicing* alternativo no câncer

A versão GRCh38.p12 do catálogo do GENCODE nosso genoma consta de 60603 genes e o transcriptoma de 226882 transcritos diferentes (Frankish et al. 2019), esses dados demonstram que virtualmente, em média, cada gene pode expressar 4 (quatro) transcritos distintos, sendo uma característica intrínseca os mecanismos de aumento de diversidade da expressão gênica da célula. Desses o mais relevante e prevalente para praticamente todos os transcritos sintetizados pelo complexo da polimerase II é o processo de *splicing* alternativo, o qual é responsável por gerar, a partir do mesmo pré-mRNA, transcritos maduros distintos (Mironov, Fickett, and Gelfand 1999; Brett et al. 2000; Croft et al. 2000).

O *splicing* alternativo é um mecanismo essencial, co-transcricional e finamente regulado que leva a expansão do repertório gênico de funções, ao produzir diferentes combinações das regiões de *splicing* em um pré-mRNA são gerados transcritos estruturalmente e funcionalmente diferentes, mais de 95% dos genes humanos sofrem *splicing* de maneira dependente do estágio de desenvolvimento, do tecido específico ou de estímulos por vias de transdução de sinal (Gilbert 1978; Yan Wang et al. 2015). O *splicing* alternativo modifica as proteínas codificadas pelos mRNAs, o que possui efeitos funcionais profundos. Dados da literatura mostram que dessas isoformas de proteínas mostrou que o *splicing* alternativo acaba por ser uma forma de regulação de ligação proteína-proteína, proteína-ácidos nucleicos, proteína-membrana, bem como são capazes de regular também a localização de proteínas, suas propriedades enzimáticas e sua interação com ligantes pela consequente supressão, adição

ou modificação de domínios proteicos, apesar de, na maioria dos casos, as alterações na estrutura primária causadas pelas AS serem pequenas (Yan Wang et al. 2015; Marinov et al. 2014). Os AS mais comuns são descritas na figura 3, “*exon skipping*” neste esquema é um caso de “*cassete exon*”, cujo éxon é desconsiderado. As retenções de íntrons em transcritos humanos estão posicionados principalmente em regiões UTRs e foram associadas com sítios de splicing fracos, tamanho de íntrons curtos e regulação de elementos regulatórios em cis. A seleção alternativa de sítios 3’ ou 5’ dentro de éxons podem levar a pequenas mudanças na sequência codificadora, e um outro caso que aumenta a complexidade é o de éxons alternativos mutuamente exclusivos (E. Kim, Magen, and Ast 2007; Katz et al. 2015; Galante et al. 2004; Sakabe and de Souza 2007).

Figura 3 – *Splicing* constitutivos e alternativos.

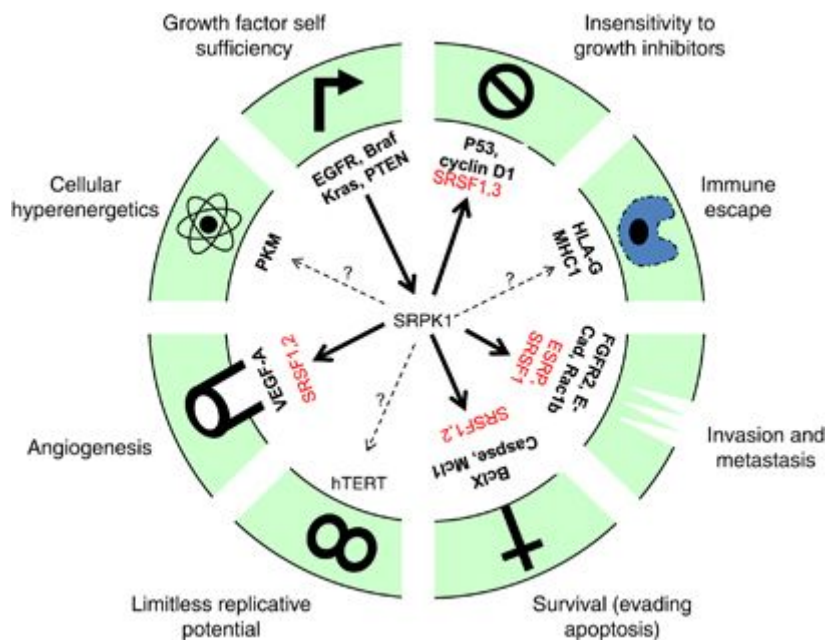


Azul claro: exons constitutivamente presentes no mRNA processado; azul-escuro: exons alternativos que podem ser incluídos ou não no mRNA maduro. Fonte (Dvinge et al. 2016)

Funcionalmente, no entanto, é bem estabelecido que o processamento AS contribui para a diferenciação celular, aquisição, manutenção de identidade tecidual ao longo do desenvolvimento do organismo. Além disso existe uma capacidade plástica regulada das células de ativar um repertório de mudanças nos programas de *splicing*, essas mudanças podem ter efeitos sobre a proliferação celular e sobrevivência celular (E. T. Wang et al. 2008; Baralle and Giudice 2017).

Algumas alterações no mecanismo de *splicing* estão envolvidas com doenças, como por exemplo mutações em elementos *cis* reguladores da expressão de RNAs, desregulações nos elementos *trans* componentes do maquinário do spliceossomo (Scotti and Swanson 2016). Já é bem estabelecido que isso se aplica ao câncer, inclusive que alguns fatores de *splicing* atuam para carcinogênese e outros como supressores de tumor (Figura 4). De forma regulada um subconjunto de fatores de *splicing* são alvos recorrentes de mutações pontuais específicas e possuem expressão desregulada no câncer (Dvinge et al. 2016).

Figura 4 – *Hallmarks* do *splicing* alternativo em câncer.



O câncer tem reguladores chave que sofrem *splicing* alternativo (AS). Os reguladores dessas formas de *splicing* são frequentemente compartilhados entre tumores, como por exemplo o SRSF1 envolvido no AS de mRNAs implicados em fenótipos tumorais como P53, FGFR2, Caspase e VEGF-A. Fonte (Oltean and Bates 2014)

São bem descritos o AS do pré-mRNA de EGFR, por produzir várias isoformas, algumas das quais são constitutivamente ativas, levando a tumorigenicidade, migração e invasão, Receptor do Fator de Crescimento Epidérmico (EGFR); fator de crescimento endotelial vascular (VEGFR), cuja alternativa As características do splicing variadas, resultam no aumento da progressão tumoral e redução da resposta à terapia (GUILLAUDEAU et al. 2012; Padfield, Ellis, and Kurian 2015). O receptor de insulina (IN-R) sofre AS sendo capaz de produzir duas isoformas: a isoforma IN-RB de comprimento total e a isoforma IN-RA que sofre *skipping* do exon 11. A isoforma IN-RB codifica um receptor com alta afinidade pela insulina, enquanto a isoforma IN-RA possui alta afinidade pelos hormônios de crescimento insulina e IGF2, isso potencializa o desenvolvimento de fenótipos de crescimento, a proliferação e a angiogênese. IN-RA é mais expressa em certos tipos de câncer como hepatocarcinoma e osteosarcoma (Khurshid 2019). Em PDAC é descrito que AS em contexto tumoral promove sobre o gene da piruvato quinase (PKM) a mudança do eixo da PKM1 para a isoforma PKM2 relacionada ao câncer, havendo uma tumorigênica que confere menor sensibilidade a gencitabina e o estresse genotóxico induzido pela cisplatina, induzindo quimiorresistência (Calabretta et al. 2016).

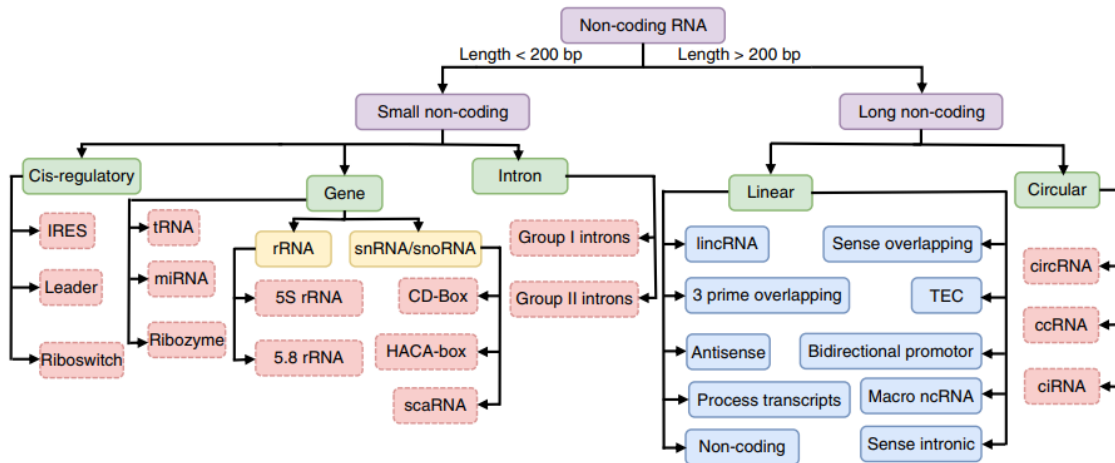
A existência de um conjunto de AS específicas para cada contexto celular tem despertado o interesse para uma abordagem como AS sendo uma classe dos Antígenos Específico de Tumores (TSAs), constituindo uma poderosa abordagem para a medicina de precisão e clínica translacional (Smith et al. 2019).

1.9. RNAs longos não codificadores (lncRNAs)

Os recentes avanços na exploração do transcriptoma, sobretudo com o advento das tecnologias NGS revelaram um comportamento do genoma humano de ter sua maior parte transcrito a RNA, cerca de 75%, e sendo que desses apenas 2% desses transcritos seguem o fluxo de informação de codificar proteínas (Djebali et al. 2012). O restante do transcriptoma é composto pela classe dos RNAs não-codificadores (ncRNA), os quais podem ainda esses ainda

podem ser subdivididos pelo tamanho do transcrito e funções biológicas (Amin, McGrath, and Chen 2019), como está descrito na figura 5.

Figura 5 – Classificação de ncRNAs. Uma proposta de divisão em classes e subclasses em eucariotos.



Fonte (Amin, McGrath, and Chen 2019)

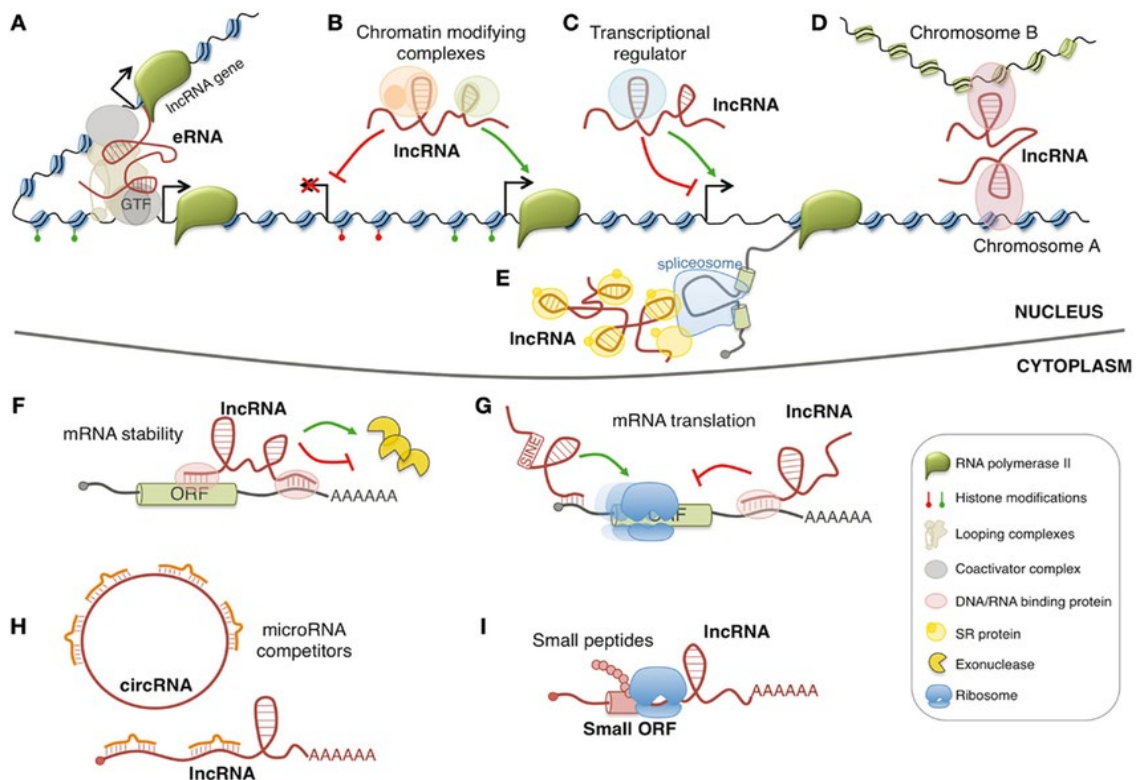
Ou seja, os ncRNAs são majoritariamente divididos em dois grandes grupos pelo tamanho: os maiores de 200nt são denominados RNAs não-codificadores longos (lncRNAs) e aqueles menores que 200nt são denominados RNAs não-codificadores pequenos (sRNAs). Os sRNAs são um grupo heterogêneo que inclui microRNAs (miRNAs), *small nucleolar RNAs* (snoRNAs) e os piwiRNAs (piRNAs) (Ma, Bajic, and Zhang 2013). Os lncRNAs por outro lado podem ser subdivididos em RNAs circulares e RNAs lineares, esses por sua vez podem ser ainda subdivididos segundo a sua localização genômica, sendo os mais relevantes os lncRNAs intergênicos (lincRNA), o qual é transcrito por um *locus* não anotado localizado entre dois genes anotados conhecidos; o lncRNAs senso intrônico, o qual é transcrito completamente dentro de um íntron de outro gene) e o lncRNA antisense, o qual é transcrito na fita oposta de outro gene conhecido, podendo haver alguma sobreposição com o locus desse gene) (Hombach and Kretz 2016; Amin, McGrath, and Chen 2019).

Entre as diversas classes de ncRNAs, os lncRNAs tem ocupado lugar de relevância no cenário atual por terem sido descritos como envolvidos no controle da expressão gênica e de processos celulares fundamentais (Figura 5), sendo postulado que possuam um papel importante na homeostase celular (Reis

and Verjovski-Almeida 2012a; Morlando, Ballarino, and Fatica 2015). Por outro lado, apesar dos esforços e acúmulo de novas evidências a respeito da função global, e na elucidação pontual da função de diferentes lncRNAs, a função da maioria dos lncRNAs permanece desconhecida (Martin and Chang 2012). Como apresentado no item 1.4, as frações de lncRNAs e mRNAs codificadores de proteína anotados no genoma são semelhantes, entre 30-40% porém apenas uma pequena fração dos lncRNAs já foram caracterizados funcionalmente (Frankish et al. 2019).

De modo geral pode-se descrever que os genes lncRNAs tem baixos níveis de conservação evolutiva comparados aos genes de mRNAs, tem menores níveis médios de expressão que os mRNAs e que a expressão de lncRNAs ocorrem de forma tecido específica, num mesmo organismo (Djebali et al. 2012). Uma vez que os lncRNAs são uma classificação arbitrária (um transcrito que não é mRNA e tem um tamanho maior que 200nt) não é surpreendente que nesse conjunto sejam detectados como envolvidos em diversos processos distintos os quais pode-se agrupar pelas funções como demonstrado na figura 6.

Figura 6 – Modelos de funções atribuídas a lncRNAs.



(A) lncRNAs nucleares podem regular a transcrição atuando como RNA enhancer (eRNA); (B) recrutando complexos modificadores da cromatina; (C) regulando a atividade dos fatores de transcrição; (D) atuando na conformação espacial dos cromossomos; (E) influenciando o spliceossomo. Os lncRNAs citoplasmáticos podem regular a expressão de mRNA: (F) regulando a estabilidade do mRNA; (G) estabilizando o complexo de tradução de mRNA; (H) competindo pela ligação de microRNA (H). (I) lncRNAs contêm pequenas ORFs que podem ser traduzidos em pequenos peptídeos ativos biológicos. Fonte (Morlando, Ballarino, and Fatica 2015)

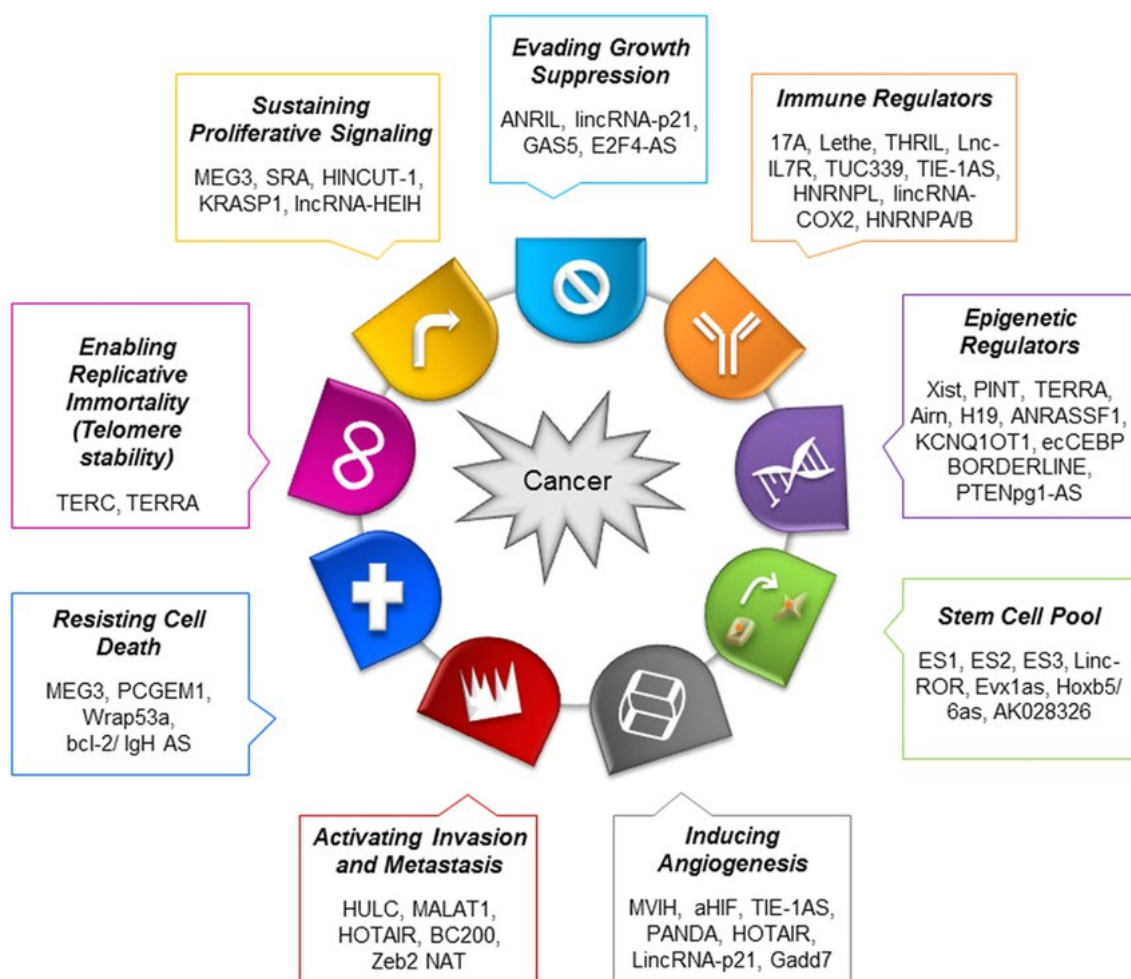
Os exemplos mais relevantes desses processos, são: a regulação da atividade transcricional, em *cis* ou em *trans*, recrutando fatores de transcrição ou complexos modificadores de cromatina (Tsai et al. 2010); a modulação do *splicing*, alterando para um processo alternativo (Gonzalez et al. 2015); a regulação dos níveis citoplasmáticos de outros RNAs pelo controle das velocidades de tradução ou degradação de mRNAs (Yoon et al. 2012; Chenguang Gong 2011). Uma dessas atuações, possivelmente a mais clássica, se dá pela atuação como esponjas de miRNAs que alvejam outros RNAs (Du et al. 2016; Shan et al. 2018). A outra função dos lncRNAs é de *enhancers* (Figura 4 - e), sendo inclusive denominados como RNA *enhancer* (eRNAs), transcritos dessa classe estão envolvidos no mecanismo de regular a expressão gênica através de sua ligação com fatores de transcrição (Jeong et al. 2016; Xue et al.

2016; Lopez-Pajares et al. 2015), alguns mais especificamente com a formação de um loop que aproxima as regiões promotoras e *enhancer* (Katie Schaukowitch et al. 2014; K. Schaukowitch and Kim 2014).

1.10. lncRNAs em PDAC

A expressão aberrante de lncRNAs reguladores tem sido cada vez mais documentada em diferentes tipos de câncer, onde podem mediar efeitos oncogênicos ou supressores de tumor pela subversão dos mecanismos envolvendo lncRNAs. É crescente o ritmo de descoberta sobre lncRNAs envolvidos em vários processos que favorecem o desenvolvimento do câncer (Figura 7), em virtude de sua capacidade de interagir com DNA, RNA, proteínas ou servir de adaptadores para combinações entre essas moléculas, e seu potencial para modular muitas das características do câncer. Alterações na regulação do ciclo celular, quimioresistência, resposta imune e pluripotência são os mecanismos bem estabelecidos nos quais lncRNAs oncogênicos atuam (Huarte and Rinn 2010; Huarte 2015; Serghiou et al. 2016). A interação com complexos de remodelação da cromatina que promovem o silenciamento de genes específicos ou a modulação de proteínas do fator de *splicing* também são outros modos gerais conhecidos de regulação oncogênica por lncRNA (Reis and Verjovski-Almeida 2012a).

Figura 7 – lncRNAs associados as *hallmarks* do câncer



lncRNAs são implicados nas principais características do câncer e podem contribuir para o aparecimento e progressão do câncer. Fonte (Parasramka et al. 2016)

Através da reprogramação da expressão gênica lncRNAs são descritos como envolvidos na reprogramação metabólica em câncer, a qual permite que as células cancerosas mantenham a proliferação desregulada e suportem o microambiente hostil, como o estresse energético (H. Liu et al. 2019).

Dentre os lncRNAs implicados em câncer tem-se PTENP1 (*Phosphatase And Tensin Homolog Pseudogene 1*) o qual atua como esponja, se-ligando a miRNAs que alvejam o mRNA PTEN (*Phosphatase And Tensin Homolog*), em câncer de mama (BC) a diminuição dos níveis de expressão de PTENP1, supressor tumoral, levam a diminuição de PTEN por ser alvejado pelo miR-19b, o que está relacionado com aumento da proliferação, migração, invasão e inibição da apoptose em BC (R.-K. Li et al. 2017). Por outro lado o lncRNA SNHG7 (Small Nucleolar RNA Host Gene 7) é descrito como tumorigênico por

atuar como esponja de miR-216b, o que leva a um aumento da expressão de GALNT1, o que por sua vez está associada com o aumento da proliferação e metástase em CRC (Shan et al. 2018).

Um outro exemplo é o HOTAIR (*HOX Antisense Intergenic RNA*), envolvido com fenótipo tumoral, descrito em BC e CRC, o principal mecanismo descrito na literatura se dá pela atuação como *scaffold* com o complexo remodelador da cromatina PRC2 (*polycomb repressive complex 2*) e reprogramação do repertório gênico correlacionado com proliferação e metástase (Gupta et al. 2010; Tsai et al. 2010; Kogo et al. 2011). O lncRNA HOTTIP (*HOXA Distal Transcript Antisense RNA*) transcrito a montante do locus dos genes HOXA, associa-se fisicamente ao fator de ligação a CCCTC (CTCF) que serve como *insulator* ao organizar o HOXA coordenando a ativação transcricional desse genes (F. Wang et al. 2018). A expressão de HOTTIP é descrita como associada ao aumento da proliferação, migração, invasão e quimiorresistência em e PC e HCC, sendo classificado como pro-oncogênico, semelhantemente ao HOTAIR (Quagliata et al. 2014; Yating Cheng et al. 2015; Zhihua Li et al. 2015).

O lncRNA-p21 é descrito como sendo supressor tumoral em câncer epidermóide de esôfago (ESCC), pois o expressão de lncRNA-p21 estimula a expressão de p21 (um dos aliases para *CDKN1A*, *Cyclin Dependent Kinase Inhibitor 1A*), a recuperação da expressão de lncRNA-p21 é relacionada com inibição da proliferação celular, da migração, da invasão e da transição do ciclo celular do G1, além de promover a apoptose (Y. Zhang et al. 2019). lncRNA-p21 também é descrito por ter expressão induzida por hipóxia celular e modular efeito Warburg por meio da estabilização de HIF-1 α em contexto de câncer, sendo considerado um lncRNA essencial para a glicólise durante a hipóxia (F. Yang et al. 2014).

O LINC00092 tem expressão aberrante em câncer de ovário e é reportado como sendo envolvido na alteração da glicólise nesse câncer. A expressão de LINC00092 é estimulada por CXCL14 (*C-X-C Motif Chemokine Ligand 14*), um fator premetastático liberado CAFs. Ao interagir diretamente com a 6-fosfofructo-2-quinase / frutose-2,6-bifosfatase (PFKFB2), uma enzima glicolítica, o LINC00092 aumenta a glicólise, promove a metástase e sustenta a

função de suporte local dos CAFs dentro do microambiente tumoral (L. Zhao et al. 2017).

Apesar dos exemplos supracitados, é importante notar, dada a infância dos estudos sobre lncRNAs, que mecanismos de ação desses permaneçam ocultos (Reis and Verjovski-Almeida 2012b; Collisson et al. 2019). Ao encontro dessa lacuna veem os estudos em larga escala de lncRNA no transcriptoma. No câncer pancreático três abordagens do transcriptoma por microarranjo exploraram os potenciais lncRNAs cuja expressão fosse relacionada com o fenótipo tumoral comparando-se amostras pareadas de PDAC (T) e amostras não-tumorais pancreáticas adjacentes (N). Em 2011, *Tahira et al.* analisando 38 amostras pareadas encontrou lncRNAs DE correlacionados com metástases em PDAC, além de constatar que esses eram enriquecidos nas vias de genes da MAPK (Tahira et al. 2011a). *Li et al.*, em 2014, encontrou 1881 lncRNAs com expressão aumentada e 3369 lncRNAs com expressão diminuída (*fold change* ≥ 4.0 ou ≤ 0.4 , p-valor < 0.01) em 30 amostras de PDAC e seus respectivos controles, propondo que baixos níveis do lncRNA BC008363 (LINC00491) se correlacionam com um bom prognóstico em casos de PDAC (Jiao Li et al. 2014). A partir de 30586 sondas para lncRNA *Zhou et al.* interrogou o transcriptoma de PDAC em 11 amostras de tumor e tecido não tumoral pareadas, detectando um conjunto de 27279 lncRNAs em PDAC, dos quais 2331 tinham expressão aumentada e 1641 tinham expressão diminuída ($|fold\ change| \geq 2$, p-valor < 0.05) e que desses a expressão de HOTAIRM1 (HOXA *Transcript Antisense* RNA, *Myeloid-Specific 1*) tem potencial biomarcador para malignidade (ZHOU et al. 2016).

Com a difusão das tecnologias NGS, Song et al. interrogou o transcriptoma do PDAC por RNA-seq, comparando dois grupos de amostras, sendo 178 amostras de tumor e 4 amostras de tecido não tumoral. Analisando-se as DE chegou ao conjunto de 5 lncRNAs (C9orf139, MIR600HG, RP5-965G21.4, RP11-436K8.1 e CTC-327F10.4) com potencial prognóstico (J. Song et al. 2018).

A robustez da abordagem por RNA-seq permitiu que esforços fossem somados para investigar o transcriptoma do câncer, incluindo-se lncRNAs, por projetos colaborativos. Um exemplo é do banco envolvendo dados analisados de 6503 bibliotecas tumorais de RNA-seq, denominado miTranscriptome, o

catálogo produzido por esse projeto contém mais de 91.000 loci gênicos, a maioria dos quais são anteriormente RNAs não caracterizados (Iyer et al. 2015). A análise da expressão gênica desses transcritos revelou numerosos RNAs com expressão restrita a tecidos tumorais e específicos para diferentes linhagens/tecidos, aumentando as descobertas e o potencial uso prognóstico e diagnóstico dos lncRNAs como biomarcadores (Reis and Verjovski-Almeida 2012b).

1.11. Redes de co-expressão gênica e WCGNA

As tecnologias de nova geração, por tratarem de análises em larga escala, têm grande aporte nas ciências *ômicas*, no contexto de biologia molecular temos principalmente a transcriptômica e a genômica. Ambas possuem uma característica intrínseca de gerarem e serem compostas por grandes conjuntos de dados, "*big data*". Com isso abordagens populacionais são propostas para analisar esses conjuntos de dados multivariados, buscando-se estabelecer alguma correlação capaz de agrupar a população de dados por alguma característica de interesse. Para atender tal necessidade, uma abordagem oriunda da biologia de sistemas, a análise de redes de correlação é largamente utilizada na análise de grandes conjuntos de dados (DiLeo et al. 2011). A análise por rede de co-expressão gênica (RCEX) mensura a taxa de correlação entre todos os genes de forma a construir a complexidade de uma rede de transcrição celular. Chama-se "co-expressão" a semelhança de padrões de comportamento de expressão gênica em uma dada variedade de condições experimentais (Aoki, Ogata, and Shibata 2007). Uma das abordagens para essa análise dá-se pela identificação de propriedades em torno das quais podem ser agrupados um subconjuntos de genes-chave de determinados processos biológicos ou vias, até mesmo pode-se ter uma abordagem cuja não seja constituídos subconjuntos por características selecionadas mas apenas a constituição de subconjuntos apenas pelos próprios valores das redes (DiLeo et al. 2011).

O estabelecimento de uma rede de co-expressão gênica trata-se de uma aplicação direta do princípio de RCEX, na qual os genes, ou transcritos, são nós, e os nós são quantitativamente conectados se os outros genes, ou

transcritos, correspondentes são significativamente co-expressos nas amostras. Em linhas gerais pode-se descrever que cada rede de co-expressão é estabelecida como a uma matriz de adjacência (A), a qual codifica a força de conexão entre dois nódulos, nesse caso dois genes ou transcritos. Ao se definir a matriz A podem ser utilizadas diferentes funções, com parâmetros estatísticos ou biológicos. Essa matriz A é, por sua vez analisada para quantificar a dissimilaridade, retratada como distância, entre os nódulos. Esta medida de dissimilaridade é empregada na formação de *clusters*, os quais são associados e denominados módulos. Esses e seus genes altamente conectados podem ser correlacionados com informações externas do gene ou da amostra (B. Zhang and Horvath 2005).

No contexto de transcriptômica a identificação em subconjuntos de genes ou transcritos em estruturas modulares é o objetivo principal numa análise de RCEX, partindo-se da premissa que os módulos representam agrupamentos de genes cujas funções podem ser correlacionadas pelas interações (Aoki, Ogata, and Shibata 2007). O processamento de dados como expressão diferencial de milhares de genes em uma RCEX com número definido de módulos reduz a complexidade dos dados o que acaba dando maior poder estatístico para posteriores análises (DiLeo et al. 2011).

O WGCNA (*Weighted Correlation Network Analysis*), um pacote do software R, contém um conjunto de funções baseadas em análise de correlação de redes, o qual descreve e permite a visualização de redes de grandes conjuntos de dados, sejam tais dados: estimativas de expressão de genes, concentrações de metabólitos ou outros dados fenotípicos. As funções do WGCNA incluem: construção de redes; detecção de módulos; seleção de módulos e genes; cálculos de propriedades topológicas; simulação de dados; visualização; e, interação com pacotes externos. Dessa forma, o pacote pode ser utilizado para explorar a estrutura de módulos dentro de uma RCEX, mensurar a relação entre genes e módulos (*module membership*), explorar a relação entre os módulos, ou mesmo, para ranquear genes ou módulos em relação aos tratamentos estudados (Langfelder and Horvath 2008). Na literatura encontram-se estudos que utilizaram a análise de RCEX gênica, por meio do WGCNA, para avaliar características complexas em diversas espécies, como camundongos (Fuller et al. 2007), em seres humanos (Heiland et al. 2016)

As análises de RCEX com WCGNA foram utilizadas em contexto de câncer para integrar dados moleculares e clínicos (M. Sun et al. 2015). Em 2016, *Heiland et al.* utilizou a abordagem para glioblastoma multiforme (GBM) análise de RCEX de *genome-wide* foi utilizada para uma análise integrativa de imagens e dados genéticos (Heiland et al. 2016). Para lncRNAs a análise da RCEX gênica pode propor correlações biologicamente funcionais e identificar suas associações de câncer em potencial, abordagem “*guilt-by-association*”. O grupo de *Tye et al.* avaliou, a partir de dados de RNA-seq, a regulação da osteogênese por lncRNAs e envolvimento em comprometimento e diferenciação celular (Tye et al. 2018). A partir da integração de integração de dados de 131 diferentes estudos de microarranjo em um único conjunto de dados expressão que teve 2968 perfis de expressão de genes, análises de RCEX e WCGNA resultou em 37 módulos de co-expressão, desses 3 módulos biologicamente interessantes foram analisados em profundidade, e potenciais funções de lncRNAs não caracterizados em leucemia e neuroblastoma foram anotadas (Cogill and Wang 2014). Schwarzer et al. avaliou lncRNAs envolvidos em leucemia e encontrou assinaturas de lncRNAs específicas, como o LINC00173 em granulócitos, além de identificar reguladores críticos envolvidos na homeostase do sangue (Schwarzer et al. 2017).

Em 2018, *Giulietti et al.* partindo de dados públicos de expressão de microarranjo, contendo 55 amostras N e 75 T avaliaram DE em análises de RCEX por WCGNA, identificando A2M-AS1, LINC01133, LINC00205 e TSPOAP1-AS1 como biomarcadores prognósticos para PDAC (Giulietti et al. 2018).

2. Materiais e Métodos

2.1. Amostras de pacientes e xenotumores

Amostras de PDAC e tecidos não malignos adjacentes do mesmo paciente, amostras pareadas, foram coletados após a ressecção cirúrgica e imediatamente imersos e congelados em nitrogênio líquido. Tais amostras compõem uma coleção de tumores de um banco mantido pelo Serviço de Patologia do Hospital A. C. Camargo. Todas as amostras de PDAC foram

classificadas histopatologicamente de acordo com o sistema TNM, recomendado pela Organização Mundial da Saúde (Scarpa et al. 2010; Kleihues and Sobin 2000), e as amostras adjacentes foram avaliadas histopatologicamente e constatadas a ausência de neoplasias.

O banco dispunha de uma coleção de amostras de PDACs, a coleção incluía tumores com os seguintes estadiamentos: estágio I, 5 casos; estágio II, 21 casos; estágio III, 0; estágio IV, 4 casos. Para aumentar o poder estatístico da análise, foram selecionadas apenas as amostras com características histopatológicas e clínicas PDAC (T). Um outro critério de inclusão adotado foi o baixo estadiamento (I e II), correspondendo às etapas em que os tumores são mais frequentemente operáveis. Além do estadiamento, outro critério de inclusão foi a existência do fragmento de tecido não tumoral adjacente (N) pareado. Pelos critérios descritos foram incluídas 28 amostras nesse estudo. 14 amostras de tumor (T) e os respectivos 14 pares não tumorais (N) (Tabela1). O estudo foi aprovado pelo Comitê de Ética da Instituição (CAAE: 15059213.0.0000.5432).

Também foram utilizados nesse estudo amostras pancreáticas de xenoenxerto derivado de paciente (PDX) de nosso laboratório, desenvolvidos pelo Dr. Luís Bruno. Essas amostras de PDX foram obtidas implantando-se fragmentos tumorais de PDAC ressectados em camundongos imunossuprimidos (Moraes 2019).

Tabela 1 – Dados clínicos referentes as amostras de tumores.

Amostra Tumoral	Amostra Normal	Sobrevida (dias)	Estatus	Sexo	Idade	Classificação anatomopatológica (H&E)
<i>Pan02T</i>	<i>Pan02N</i>	397	Óbito	M	65	80% carcinoma adenoescamoso, 20% bordeline intestino delgado.
<i>Pan04T</i>	<i>Pan04N</i>	405	Vivo	M	62	30% IPMN de baixo grau, borda não neoplásica removida.
<i>Pan06T</i>	<i>Pan06N</i>	936	Vivo	F	61	90% PDAC moderadamente diferenciado, 5% de lesão <i>in situ</i> não separável, 10% de borda não neoplásica removida.
<i>Pan12T</i>	<i>Pan12N</i>	369	Óbito	M	81	95% PDAC moderadamente diferenciado, 5% neoplasia intra-ductal de alto grau inseparável.
<i>Pan25T</i>	<i>Pan25N</i>	3291	Óbito	M	67	80% PDAC, 20% de borda de não neoplásica removida.

<i>Pan29T</i>	<i>Pan29N</i>	448	Óbito	F	75	100% PDAC.
<i>Pan30T</i>	<i>Pan30N</i>	115	Óbito	M	77	100% PDAC.
<i>Pan35T</i>	<i>Pan35N</i>	80	Vivo	M	77	80% PDAC, 20% de tecido pancreático endócrino inseparável.
<i>Pan38T</i>	<i>Pan38N</i>	658	Vivo	M	53	80% PDAC, 20% de tecido pancreático endócrino inseparável.
<i>Pan39T</i>	<i>Pan39N</i>	1608	Vivo	F	46	70% PDAC, 30% de tecido pancreático não neoplásico inseparável.
<i>Pan44T</i>	<i>Pan44N</i>	1077	Vivo	M	80	100% IPMN.
<i>Pan47T</i>	<i>Pan47N</i>	1170	Vivo	F	67	100% PDAC.
<i>Pan50T</i>	<i>Pan50N</i>	531	Óbito	F	65	90% PDAC, 10% de borda de parede de duodeno inseparável.
<i>Pan56T</i>	<i>Pan56N</i>	1287	Vivo	F	75	100% neoplasia intra-epitelial pancreática.
<i>Pan58T</i>	<i>Pan58N</i>	252	Óbito	M	73	80% PDAC, 20% de mucosa normal inseparável.

2.2. Linhagens celulares, linhagem celular recém-derivadas de PDX e cultivo de tumoresferas

As linhagens de células de PDAC foram adquiridas da American Tissue Culture Collection (ATCC): Capan-1 (ATCC® HTB-79), AsPC-1 (ATCC® CRL-1682), MIA PaCa-2 (ATCC® CRL-1420), BxPC-3 (ATCC® CRL-1687), PANC-1 (ATCC® CRL-1469). A linhagem de fibroblastos pulmonares imortalizados MRC-5 foi fornecida gentilmente pelo professor Carlos Frederico Martins Menck do Laboratório de Reparo de DNA.

As células foram propagadas segundo orientações da empresa fornecedora: as linhagens de Capan-1, MIA PaCa-2, PANC-1, MRC-5 e PDX em meio DMEM, já as linhagens de células AsPC-1 e BxPC-3 foram cultivadas em meio RPMI 1640. Todos os cultivos foram suplementados com 10% de soro bovino fetal, excetuando-se Capan-1, que teve 20% de suplementação de soro fetal bovino; e todos cultivos receberam 1% de penicilina-estreptomicina. As células foram cultivadas numa incubadora umidificada em atmosfera de 5% de

CO₂ a 37 °C. As proliferações celulares foram realizadas em colaboração com a especialista Bianca Dazzani do nosso laboratório.

Tumoresferas e células aderentes AsPC-1 foram cultivadas em nosso laboratório pelos doutorandos Thalita Bueno e Lutero Hasenkamp, orientados pela profa. Daniela Sanchez Bassères. As células foram cultivadas em placas de cultura de baixa aderência (Corning) de 6 (seis) poços com 2mL de meio DMEN F12 contendo EGF, FGF e B27 na concentração de 1µL/mL. As células foram suplementadas com 500 µL de meio fresco a cada 3 dias, e mantidas em cultura por um total de 7 dias em atmosfera úmida (95% O₂; 5% CO₂) a 37°C. Então as tumoresferas foram coletadas (12 poços formando um *pool*) e centrifugadas para a extração de RNA.

2.3. Extração de RNA

O RNA total das amostras tumorais (T) e das amostras não tumorais adjacentes (N) foram realizadas em parceria com o Serviço de Patologia do Hospital A. C. Carmago, sob a coordenação de Doutora Maria Dirlei Begnami. O RNA total das linhagens celulares e PDX utilizados foram realizados no Laboratório de Pesquisas Aplicadas em Câncer (LAPIC).

Todo o RNA foi isolado das amostras tumorais (T) e das amostras não tumorais adjacentes (N) utilizando-se o reagente *Trizol* (ThermoFisher) e purificando-se de acordo com o protocolo do fabricante. Após a coleta da fração aquosa contendo o RNA total, à fase orgânica formada foi adicionado clorofórmio (500µL para cada 1mL de Trizol) e novamente realizada a centrifugação. A fase aquosa resultante foi precipitada com isopropanol gelado e o precipitado foi lavado com solução de etanol 70% gelado e deixado secar a temperatura ambiente. Uma vez evaporado todo o etanol procedeu-se para a o tratamento com DNase I (ThermoFisher) de acordo com as informações do fabricante para eliminar contaminações por DNA da amostra. O procedimento foi realizado adicionando-se 1µL de tampão 10x da enzima (100mM Tris-HCl, pH7,5, 25 mM MgCl₂, 1mM CaCl₂), 8 µL de água DEPC e 1µL de enzima DNase I. As amostras foram incubadas por 30 min a 37°C, em seguida foi feita a parada da inibição da DNase I por adição de 1µL de solução de EDTA (25mM) e incubação por 10 min a 65°C. As amostras então foram armazenadas em ultrafreezer a -80°C ou

utilizadas imediatamente. A qualidade do RNA utilizado para a preparação da biblioteca de cDNA foi verificada utilizando o chip de análise de RNA nano 6000 (Agilent, Cat # 5067-4627) no instrumento BioAnalyzer 2000 (Agilent Technologies) para assegurar um valor de integridade de RNA (RIN).

2.4. Preparo de bibliotecas e sequenciamento

As bibliotecas de RNA-Seq foram geradas utilizando-se para o preparo o reagente Illumina TruSeq Stranded Total RNA LT com Ribo-Zero Gold (Cat # RS-122-2001) de acordo com o protocolo padrão do fabricante (Part n1 15031048 Revisado em abril de 2013). A primeira inicial é a de depleção de RNA ribossômico, que foi realizada usando-se 1 µg de RNA de cada amostra. O objetivo foi produzir bibliotecas cujo tamanho médio final estivesse na faixa de 120 a 200 pb. O RNA foi fragmentado quimicamente e o cDNA foi então sintetizado utilizando-se a enzima Transcriptase Reversa *SuperScript II* (Invitrogen, (Cat # 18064) e iniciadores randômicos. Uma vez que o sequenciamento é fita-específica, o cDNA de fita simples resultante foi convertido em DNA de de dupla fita, pela estratégia da marcação da fita complementar, a qual foi sintetizada presença do nucleotídeo de dUTP para prevenir a subsequente amplificação da segunda cadeia. Segue-se então para adição específica de uma adenina a extremidade 3' e subsequente ligação do adaptador nessa extremidade. As bibliotecas foram então submetidas a 15 ciclos de PCR para produzir bibliotecas de RNA-seq prontas para seqüenciamento. A qualidade das bibliotecas de RNA-seq geradas foi verificada por eletroforese capilar (Agilent High Kit de Análise de DNA de Sensibilidade, Cat # 5067-4626) no instrumento BioAnalyzer 2000 (Agilent Technologies). A quantificação da biblioteca foi realizada por qPCR usando o KAPA Library Quantification Kit - Illumina / Universal (KAPA Biosystems, Cat # KK4824). O sequenciamento ("Paired-end") de baixa cobertura na plataforma Illumina MiSeq foi realizado para confirmar a qualidade das bibliotecas geradas e, em seguida, foi realizado o sequenciamento de maior cobertura em um sistema HiSeq 2500 com os reagentes TruSeq SBS Kit v3 - HS (200 ciclos) (Illumina, Cat # FC-401-3001), para gerar leituras de 100 pares de bases. A profundidade média da amostra de cobertura foi de 3,4 Gb.

O sequenciamento propriamente dito, “*Paired-end*”, foi realizado utilizando kit v3 de 150 ciclos (2 x 75 nt) na plataforma Illumina HiSeq 1500 do Instituto Butantan, no modo “*High Output Run*”.

2.5. Análises bioinformáticas do RNAseq

As análises bioinformáticas a foram realizadas em colaboração interna no nosso laboratório pelo então aluno de doutorado Julio Sosa como parte da sua pesquisa do doutorado (Sosa 2019a).

O processamento de dados de RNA-Seq total e fita-específico gerados a partir das 14 amostras pareadas (item 2.1, tabela 1) a anotação de novos transcritos e novas formas de *splicing*, análises de expressão diferencial entre amostras T e N, enriquecimento de vias. Para isso os dados provenientes do sequenciamento NGS passou pela demultiplexação dos arquivos de saída, gerando cada uma das 28 bibliotecas referentes a cada amostra. Em seguida foram feitas as análises de controle de qualidade dos fragmentos, sequências correspondentes (*reads*), gerados pelo sequenciamento utilizando-se o software FastQC. Os *reads* resultantes foram então mapeados e alinhados contra as sequências no genoma humano GRCh37/hg19 de referência utilizando-se o programa TopHat v2.0.11 (D. Kim et al. 2013). Por sua vez a montagem do transcriptoma foi realizado separadamente, a montagem das bibliotecas não-tumorais (N) foi realizada separadamente das bibliotecas tumorais (T) usando o programa StringTie (Pertea et al. 2015). Cada transcrito reconstruído recebeu um código alfanumérico de identificação (ID) pelo programa (TCONS) seguindo de um número de 8 algarismos.

Como optamos pela abordagem do transcriptoma a partir de RNAseq de RNA total para investigar também a fração poliA- do transcriptoma, adotamos um segundo grupo de filtros com parâmetros estritos na tentativa de eliminar artefatos da reconstrução. O primeiro desses filtros consistiu em remover quaisquer transcritos duplicados, ou seja, que continham exatamente as mesmas coordenadas genômicas. O segundo filtro removeu possíveis pre-mRNAs, que se apresentavam como transcritos mono-exônicos de genes confirmadamente anotados como multi-exônicos na referência Gencode. Por fim

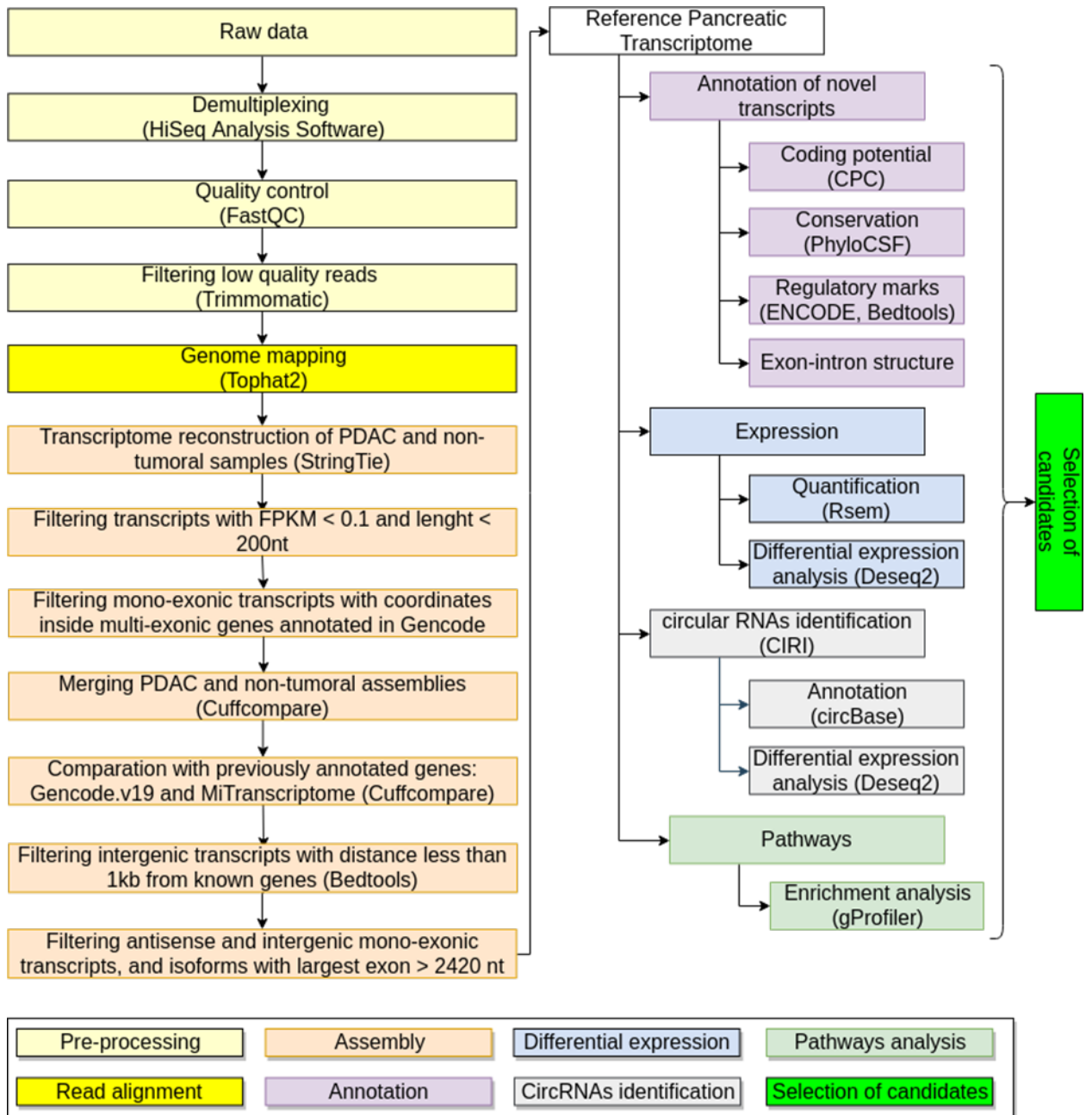
os transcritos com baixíssima expressão (FPKM menor que 0,1) e tamanho menor que 200nt também foram excluídos do conjunto.

As montagens N e T foram então combinadas e os transcritos foram então anotados e classificados em função de catálogos de referências externas o GENCODE GRCh37/hg19 e o miTranscriptome (Iyer et al. 2015) pelo programa *Cuffcompare* (Trapnell et al. 2012). Assim os transcritos além do ID conferido pelo *StringTie* também receberam as identificações correspondentes dos outros dois catálogos.

Para visualização dos transcritos reconstruídos foram gerados *tracks* customizadas cuja visualização foi possível no navegador *UCSC Genome Browser* (Kent et al. 2002; Karolchik et al. 2014), tendo como referência a montagem GRCh37/hg19 do genoma humano. As *tracks* foram sempre consultadas pelo link: <https://bit.ly/2JJ3pQB>.

Na análise de seleção de genes/transcritos investigados adotamos uma última camada de filtragem também para restringir a possibilidade de artefatos resultantes da presença de pré-mRNA das bibliotecas, optamos por não examinar transcritos novos (não anotados nos catálogos de referência do GENCODE e miTranscriptome) que eram mono-exônicos, apenas constituíram o conjunto analisado os novos transcritos intergênicos e antisensos multi-exônicos, ou seja, aqueles com evidência de *splicing*. Desse conjunto, foram excluídos aqueles transcritos classificados como novas isoformas que apresentavam éxons maiores que 2420nt e que esse não era sustentados por nenhum dos dois catálogos, pois éxons aberrantemente longos poderiam ser artefatos de montagem causados por pré-mRNA processados antes da adição de cauda poliA, e visto que 99% dos transcritos anotados no GENCODE tem éxons com tamanho menor que 2420nt. A visão global desse processo pode ser observada na figura 8.

Figura 8 – Fluxograma geral da rotina empregada na análise computacional dos dados de RNAseq de tumores de pâncreas.



As diferentes etapas do pipeline informática estão representadas por cores. Fonte: (Sosa 2019b).

Foram avaliados os níveis de expressão para cada gene e para cada transcrito reconstruído nos dados RNAseq, para a análise de expressão diferencial comparando-se as bibliotecas T e N. A análise a nível de gene contabilizou os *reads* em um único valor para uma determinada região gênica considerando-se apenas o *locus* que contém todos as regiões de DNA correspondente aos éxons. Essa análise foi realizada utilizando-se o programa

HTSeq, para alinhar os *reads* com as coordenadas de referência, e o programa RSEM que realiza a quantificação e comparação desses (Anders, Pyl, and Huber 2015; B. Li and Dewey 2011). Uma outra abordagem foi realizada para poder-se avaliar os níveis de expressão para as distintas isoformas de transcritos reconstruídos, nesse caso os *reads* foram contabilizados com parcimônia para cada isoforma específica, o que gerou valores de expressão para cada transcrito de um mesmo gene presente na montagem, para a quantificação e DE foi utilizado o programa Salmon (Patro et al. 2017), também foi feita em paralelo a mesma análise utilizando o RSEM e outra que foi utilizado o programa Kallisto (N. L. Bray et al. 2016).

O subconjunto dos genes diferencialmente expressos (DE) provenientes da análise anterior foi realizada uma análise de enriquecimento de vias. Para isso foi utilizados as ferramenta *gProfiler* (Reimand et al. 2016), a qual se utiliza informação de bancos de dados como Gene Ontology (Ashburner et al. 2000), KEGG (Kanehisa and Goto 2000) e *Reactome* (Fabregat et al. 2018).

Os transcritos novos, ou seja, aqueles reconstruídos que não estavam nos catálogos de referência (GENCODE.v19) foram analisados para identificar possíveis lncRNAs expressos em tecido pancreático, tais análises envolviam a avaliação de características como potencial codificador, conservação evolutiva, presença de marcas regulatórias e tipos de *splicing* das novas isoformas (Sosa 2019).

2.6. Análises por redes de co-expressão e WGCNA

As análises de rede de co-expressão foram realizadas em nosso laboratório pelo aluno de doutorado Diogo Vieira Pellegrina, aluno do programa Interunidades de Bioinformática O pacote do software R, WGCNA foi empregado para agrupar genes em função de o quanto são expressos em cada uma das amostras, e dentro de cada um destes grupos chamados módulos (Langfelder and Horvath 2008), e encontrar lncRNAs com maior conectividade em uma via metabólica que os mRNAs que a definem. Os dados de expressão gênica foram convertidos para TPM (transcritos por milhão) para serem usados como entrada. A matriz de correlação entre todos estes transcritos foi calculada, e a fim de criar uma matriz de adjacência livre de escala estes valores foram elevados a uma

constante, que foi definida como $\beta = 8$ (R^2 livre de escala = 0,95). Foram gerados 22 módulos, desses 2 foram selecionados pelas características biológicas e foram analisados em maior profundidade.

2.7. Análises de sobrevida

As análises de sobrevida foram realizadas em nosso laboratório pela Ester Risério Matos, aluna de doutorado do programa Inter unidades de Bioinformática.

Os níveis de expressão dos transcritos reconstruídos foram obtidos a partir de dados públicos disponíveis nos bancos de dados do ICGC PACA-AU (n= 76) e PAAD-US (n=113) (Hudson (Chairperson) et al. 2010; J. Zhang et al. 2011). Foram extraídos desses bancos os dados clínicos de *status* vital e tempo de sobrevida dos pacientes, e os dados brutos de RNAseq desses bancos foram analisados e quantificados em transcritos por milhão (TPM) a partir da montagem do nosso transcriptoma. Os filtros eliminatórios adotados foram os seguintes: amostras cujas alguma dessas informações estivessem ausentes foram removidas, bem como aquelas com sobrevida menor do que 60 dias, uma vez que poderia ser atribuída outra causa mortis que não o PDAC, como complicações pós cirúrgicos por exemplo. Além dos filtros utilizou-se de censura direita na análise, limitando-a ao período de 3 anos, amostras cuja sobrevida associada era maior que esse foi censurada, porém analisada com as demais. Para cada transcrito de interesse, os pacientes foram divididos de acordo com a mediana da expressão em TPM, originando sempre dois grupos chamados *high* e *low*. A partir desses dados foram construídas curvas de sobrevida para os dois grupos, que foram comparadas pelo teste *log-rank* (Therneau and Grambsch 2000; Therneau n.d.). Apenas foram consideradas significativas diferenças de sobrevida cujo *p*-valor < 0,05. A visualização dos gráficos foram geradas com auxílio do pacote *survminer* do R (Kassambara et al. 2019).

2.8. Síntese de DNA complementar por transcrição reversa (RT) seguida de reação em cadeia da polimerase quantitativa (qPCR)

As reações de transcrição reversa tanto para RNA proveniente de linhagens celulares, linhagem derivada de PDX, PDX e do RNA extraído das amostras tumorais e não-tumorais de pacientes foram realizadas utilizando-se o kit *SuperScript™ III First-Strand Synthesis System* (Cat# 18080051), as condições utilizadas foram as recomendadas pelo fabricante. Mais especificamente 6µL de cada RNA, em concentração normalizada para cada lote de experimentos, foi incubado a 65°C por 5min com 1µL de iniciadores (*primers*) *random hexamer* (50ng/µL), acompanhados de 1 µL de *Annealing Buffer* e água tratada com DEPC, totalizando 8µL. Após desnaturação do RNA, foram adicionados 10µL de *2X First-Strand Reaction Mix* e 2µL de *SuperScriptIII/RNase OUT Enzyme Mix* e a reação foi mantida a por 7min a 25°C para a hibridização dos *primers* e 50°C por 50min para a transcrição, seguido de 5 min a 85°C para inativação da enzima. Os 20µL da reação de RT foram diluídos com água tratada com DEPC. O DNA complementar gerado foi utilizado como substrato em reação em cadeia da polimerase quantitativa (qPCR) para uma comparação relativa dos níveis de expressão detectados. Para cada amostra e condição foram realizadas reações em triplicata de qPCR semi-quantitativo. Por reação foram utilizados 3µL de cDNA proveniente das amostras com 3µL dos *primers forward* e *reverse* (para uma concentração final de 600nM) e 6µL de *SYBR Green PCR Master Mix* (Applied Biosystems). A reação foi incubada a 95°C por 20s e 40 ciclos de 95°C por 3s, para desnaturação e 60°C por 1 minuto para o anelamento e extensão.

Para a transcrição reversa seguida de PCR quantitativo em tempo real (RT-qPCR) adotou-se a estratégia de interrogar o transcrito sempre a partir de sítios de amplificação em borda de éxons adjacentes, flanqueando de regiões de *splicing*, a fim de evitar amplificações espúrias. Para os transcritos investigados cuja a estrutura é nova, ou seja, não anotada nas referências, essa estratégia foi fundamental para se descartar a possibilidade de tratar-se de um artefato de montagem do transcriptoma, visto que as bibliotecas (item 2.4) foram construídas a partir do protocolo de RNA depletado de rRNA (S. Zhao et al. 2018).

Para isso os iniciadores foram desenhados com o auxílio das ferramentas online Primer3 4.0 (<http://primer3.ut.ee/>) e IDT Oligo Analyzer 3.1 (<https://www.idtDNA.com/>) e os parâmetros fornecidos específicos para qPCR. Preferencialmente, foram escolhidos para a validação os iniciadores mais

próximos as bordas dos éxons adjacentes, que apresentavam tamanho entre 20 a 24 pares de base (excetuando-se os primers para HMBS proveniente da literatura de trabalhos anteriores do laboratório), temperatura de hibridização entre 57 e 62°C, conteúdo de GC entre 40-70%, um máximo de três citidinas ou guanosinas consecutivas. Outro critério de eliminação para a escolha dos iniciadores foi a possibilidade do par selecionado também ter o potencial de amplificar outras regiões do genoma, e para essa finalidade foi utilizada a ferramenta *Primer BLAST* - NCBI (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>), a qual simula uma reação de PCR (*PCR in silico*) com o par de iniciadores de interesse, com base na sequência do genoma referência. Por fim o último critério foi o teste de eficiência em concentrações crescentes de cDNA, sendo que só foram incluídos pares de iniciadores que apresentaram eficiência de amplificação entre 70 a 130% (Tabela 2, Tabela 3 e Tabela 4). Os experimentos de RT-qPCR foram realizados no equipamento *Applied Biosystems® 480 7500 Real-Time PCR System* (Thermo Fisher). A normalização das expressões detectadas foi efetuada usando-se como referência os níveis detectados do gene endógeno codificador para a hidroximetilbilano sintase (*HMBS*) em todos os experimentos de RT-qPCR (Tahira et al. 2011a).

Tabela 2 – *Primers* usados na RT-qPCR dos lncRNA conhecidos DE

Gene	<i>Primer Forward</i>	<i>Primer Reverse</i>
<i>LINC01559</i>	CTCTGAAACGAAGGGCTGAC	TGCAATCCAGCAGAATGTGT
<i>LINC01133</i>	CCTTTGGATCTGGAAAGCTG	TCTTTACCTCCTCCCAACCA
<i>LINC00920</i>	ACAGCCCTGAAGCAAAACAG	GGTCTTGGAAAGTGGGTCCTT
<i>LINC01614</i>	ATGTTCTTCCTCCCAGGAT	TCGCTCTTGGTTGACATTTG
<i>MIR210HG</i>	CTGAAGTTGGGCCGAGAG	ATGGTCCTGTTGGCTGAAGA
<i>CCAT1</i>	AAAGGTGCCGAGACATGAAC	GGAGAAAACGCTTAGCCATACA
<i>LINC00675</i>	GAGGAGAAGAGACGCAGGTG	GTATTGCAGCCTCTCGTTCC
<i>LINC02577</i>	TGTCAACACGAAGGCAGAAG	ACTAGTGGCCCCAGGAAGTT
<i>LINC00842</i>	GCTCTGAATCGTGTTCTGA	TTCTTCTGCATCCCCACATC
<i>UCA1</i>	TTCCGTGAAGAGATCCACCT	CCTCCACCGTAAGAGTTACCC
<i>HMBS^a</i>	GGCAATGCGGCTGCAA	GGGTACCCACGCGAATCAC

^a Gene utilizado como normalizador endógeno.

Tabela 3 – *Primers* usados na RT-qPCR dos novos transcritos DE.

Transcrito	<i>Primer Forward</i>	<i>Primer Reverse</i>
------------	-----------------------	-----------------------

TCONS00085964 AGGTCCAGAGAGAAAAGTGACT TACATTTGGCATTGCGGTGA
 TCONS00009076 GCACGTCTGACTGTCAATCC TTCTCCTCTTCTTGTACCACAC
 TCONS00087289 GGAGAGACAGAAGAACCAAAGA ACATCCCTAGGCATCATTAG
 TCONS00006750 CTGAGGAGATAGCCCGAGAA GGACAGTGTTCCTCCAGACTT
 TCONS00059572 AGGTAAAGGAGAAGCAGGTGT TCCATAAACTTACCCAGCCA
 TCONS00036574 GGCACCTTCTGGAATCGTTC CGATGGAAGCAGCACTGTCT

Tabela 4 – Primers usados na RT-qPCR de genes codificadores avaliados.

Gene	Primer Forward	Primer Reverse
<i>B3GNT3</i>	ACGACTCCTTCTTCAACCTCAC	ACCATGTTGTCTGTGTGTGC
<i>GALNT3</i>	TCTTGGACCAGACTCGAC	TTCTAAGCAACGTGGACCAC
<i>MUC13</i>	CCCCACCACAGAAGACAATC	ACCCTTCTAAACACAGGCAAA
<i>SPIDR</i>	AGAATGCCCATCCTTTCCAG	AGGTGCTTTTCTGTAATCGTC
<i>PPP4C</i>	ATGACCTCAAAGAGCTGTTCAG	TAGAAGCCACGGTCCACAAAG
<i>ALDH1</i>	GACAAAGTAGCCTTCACAGGATCA	TCCCGGCAGCTTCTTTGA
<i>MMP2</i>	GAAGTTCTTTGGACTGCCCC	CCAGATCAGGTGTGTAGCCA
<i>MMP9</i>	CAGACCTGGGCAGATTCCA	CAAAGGCGTCGTCAATCACC

O controle da reação de qPCR para cada gene foi feito pela análise das curvas de *melting*. Essa abordagem clássica largamente utilizada faz-se comparando-se a T_m das reações completas da qPCR (todos os reagentes necessários e o cDNA substrato) com os controles sem cDNA (NTC) e controle sem transcriptase reversa (no RT). Uma vez que a T_m de cada produto de amplificação depende do seu conteúdo em guanina e citosina (G+C), comprimento e características da sequência, cada produto demonstrou uma T_m muito específica permitindo distinguir a presença de diferentes produtos daqueles esperados da PCR. Todas as qPCRs das amostras foram comparadas com os controles NTC e no RT, verificou-se que os produtos principais da qPCR não se tratavam de artefatos como ligações não específicas do corante a qualquer produto de dupla cadeia, incluindo dímeros de *primers* e produtos de amplificação não específicos, bem como não se tratava de contaminação da amostra com DNA genômico (Espy et al. 2006; Mackay, Wright, and Bonfiglioli 2008).

O nível de expressão de cada transcrito foi calculado por quantificação relativa $\Delta\Delta CT$ (Livak and Schmittgen 2001; Schmittgen and Livak 2008). Comparações entre os grupos de amostras foram representadas por *fold change*. A significância estatística da diferença de expressão foi calculada

utilizando o Teste T uni-caudal não-pareado, com correção de *Welch*, sendo apenas *p*-valores < 0,05 considerados como significativos.

2.9. PCR *end point* seguida de eletroforese

As reações de PCR *end point*, ou PCR convencional foram realizadas para se validar a existência de regiões específicas dos transcritos em análise. Os *primers* (Tabela 5) foram desenhados utilizando-se os mesmos parâmetros que os descritos no item 2.7 sendo então passível tanto de uso em PCR convencional quanto para qPCR, entretanto houve *primers* cuja natureza era incompatível com a qPCR dado o tamanho os amplicons esperados (Tabela 6). Para a execução da PCR foi utilizado o kit *GoTaq® DNA Polymerase* (cat.# M3001) e as condições utilizadas foram aquelas recomendadas pelo fabricante. Foram elas as descritas a seguir: a cada reação foi adicionada 5µL de *Reaction Buffer (5X)*; 0,5µL *dNTP Mix (10 mM)* (cat.# U1511); 6,25µL de *primer forward* (de uma solução estoque a 10 µM); 6,25µL de *primer reverse* (de uma solução estoque a 10 µM), 6,38 µL de água DEPC; 0,13µL de *Enzime GoTaq (5 U/µL)* e 0,50µL de cDNA, totalizando um volume de 25µL. Em seguida o sistema foi submetido a uma incubação a 95°C por 5min e subsequentemente 35 a 40 ciclos de 95°C por 30s, 60°C por 30s e 72°C por 40s, finalizando com uma incubação a 72°C por 7min.

Tabela 5 – Primers usados na PCR end point das novas formas de splicing DE

Gene	Primer Forward	Primer Reverse
<i>MMP14</i>	GCAGAGGGAGGGACATAGAC	GTTGGGGACGTGGTTGTTTT
<i>CAPN8</i>	ACTTTTGAATGAGGCGTTTTCC	TTTTACCCGTCCCTTCCTTCT
<i>LIF</i>	CACTGAAGACCTGGCAAACC	CGCCAAGACCTTCATTATGGG
<i>OCT3</i>	GAAATTGACCTTGGCTCCCA	CGATCCTGCCATACCTGTCT

Tabela 6 – Primers usados na PCR end point dos novos transcritos intergênicos.

Transcrito	Primer Forward	Primer Reverse
<i>TCONS00085964</i> (<i>éxons 1 e 3</i>)	GCCTTGTTTTGACACCCAGG	TACATTTGGCATTGCGGTGA

<i>TCONS00085964</i> (<i>éxons 2 e 3</i>)	AGGTCCAGAGAGAAAAGTGACT	GTTGTCCAAGCAGAGATGACT
<i>TCONS00009076</i> (<i>éxons 2 e 1</i>)	AGCTGGATATCGATTTACCCTC	TTCTCCTCTTCTTGTACCACAC
<i>TCONS00009076</i> (<i>éxons 1 e 2</i>)	GCACGTCTGACTGTCAATCC	ATGCTCCACGTGTAGGGTAC
<i>TCONS00087289</i> (<i>éxons 1 e 2</i>)	TGACTCTTGAAAACCATGGCA	TTCCTGGCATAAACAAGTCTCA

Os produtos da PCR *end point* foram submetidos a eletroforese em gel de agarose 2% (m/v) para produtos esperados entre 50 a 500bp, e gel de agarose 1,5% (m/v) para produtos maiores que 1Kb. No preparo de agarose 2% (m/v) foi misturado 2,0g de agarose pura sólida para cada 100mL de tampão TAE, o mesmo tampão foi utilizado como tampão de corrida (Hayward and Smith 1972). A mistura foi então aquecida no micro-ondas, potência máxima, por 1-3 minutos até que a agarose estivesse completamente dissolvida. Em seguida a solução de agarose esfriou em temperatura ambiente até cerca de 40°C (ponto no qual era possível segurar confortavelmente manter o frasco na mão. Então foi adicionado o agente intercalante Brometo de Etídio (EtBr) na solução de agarose antes do início da solidificação para garantir a homogeneidade da mistura. O EtBr foi adicionado para uma concentração final de aproximadamente 0.5 µg/mL, isso correspondeu a 3µL da solução estoque (Aaij and Borst 1972). 6µL cada amostra foi aplicada em cada poço do gel e o sistema foi submetido a uma diferença de potencial de 80V (8V/cm²) até que os marcadores de baixo peso molecular do tampão de corrida percorressem toda a extensão do gel (Sambrook, Fritsch, and Maniatis 1989). Todos os resultados foram visualizados por coração brometo de etídeo adicionado ao gel e revelados em luz UV em sistema ChemiDoc™ XRS+ da Bio-Rad.

2.10. Silenciamento transiente por siRNA

Foram sintetizados siRNAs para 5 lncRNAs alvos em estudo (LINC01559, LINC01133, CCAT1, LINC00920 e UCA1), optou-se por ter-se pelo sistema *Custom SMARTpool ON-TARGETplus* (Dharmacon), o qual é constituído por uma mistura de 4 sequências de siRNAs para o mesmo alvo, cada *pool* recebeu neste estudo respectivamente a seguinte denominação:

siLINC01559, siLINC01133, siCCAT1, siLINC00920 e siUCA1. A escolha dessa estratégia em detrimento de utilizar siRNAs de sequência deu-se por ser notória a maior eficiência de silenciamento e menor efeitos *off-target*. Como controle foi sintetizado um *pool* siRNA scramble, denominado neste estudo siCTRL.

Como veículo para transfecção das células com siRNA utilizou-se a *Lipofectamine™ 3000 Reagent* (ThermoFischer), para isso seguiu-se as orientações do fabricante e padronizou-se as seguintes condições de transfecção: células de AsPC-1 de baixa passagem em cultivo foram semeadas em quantidade suficiente para o estabelecimento de confluência entre 70 a 90%. Para placas de 6 poços com área de crescimento de 9.5cm² (Corning® Costar® - cat#CLS3506) foram semeadas 10⁵ células, essas foram cultivadas em 2mL de meio RPMI com 10% SFB (conforme item 2.2) por 48h. Para as placas de 100mm com área de crescimento de 55cm² (Corning® Costar® - cat# CLS3295) foram semeadas 2x10⁶ células, essas foram cultivadas em 10mL de meio RPMI com 10% SFB por 48h. O preparo de todas as soluções se deu em ambiente estéril, capela de fluxo laminar. Um volume de 2µL *Lipofectamine™ 3000 Reagent* foi gentilmente diluída em 98µL de meio RPMI sem SFB e sem antibiótico e incubado por 20min. Enquanto isso 2,5µL de siRNA foram adicionados em 22,5µL de tampão para siRNA, a mistura foi incubada por 5min e em seguida misturada a 75µL de meio RPMI sem SFB e sem antibiótico. A solução de lipofectamina e a solução de siRNA foram misturadas na proporção 1:1 e adicionado sobre as células. Em seguida eram adicionados 800µL de meio RPMI com SFB e sem antibiótico para um volume final de 2mL por poço (placa de 6 poços). Os silenciamentos realizados em placas de 100mm foram feitos com os reagentes nas mesmas proporções para um volume final de 10mL. Em ambos os casos, uma vez adicionados os reagentes todas as células foram incubadas por 48h a 37°C. Todos os silenciamentos foram validados por RT-qPCR.

2.11. Ensaio de proliferação e migração por imageamento ao vivo.

Este ensaio foi realizado pela especialista Bianca Dazzani em nosso laboratório.

Células AsPC-1 tratadas com siRNA e os respectivos controles foram semeadas na quantidade de 2×10^4 e foram cultivadas para baixa densidade em poço de 35mm numa placa de 24 poços. Cada poço foi monitorado por microscópio *zenCELL owl* 24 canais durante 8 dias, para isso foram geradas imagens fotográficas de cada poço em intervalos de 20min. A partir das imagens de cada amostra foi feita a análise e comparação das taxas de crescimento celular, sempre entre as condições de silenciamentos de lncRNAs e a condição controle (Kleinhans et al. 2018). As análises estatísticas foram feitas utilizando-se ANOVA.

Para a análise de migração celular foi utilizado um período de 6h e as o programa *ImageJ* foi utilizado para análise de migração de 15 células em cada réplica e determinação das respectivas velocidades médias (Schneider, Rasband, and Eliceiri 2012; T. J. Collins 2007).

2.12. Ensaio de migração e invasão *Transwell*[®]

Células AsPC-1 após transfectadas com siRNA foram cultivadas por 48h em meio RPMI-1640 com 10%SFB, em seguida semeadas na quantidade de 5×10^5 células em placas 6 poços, e cultivadas por mais 24h em meio RPMI-1640 com 10%SFB. 4. Após a incubação cada cultivo foi lavado com PBS e incubado por mais 24h em meio RPMI-1640 sem soro.

As células foram removidas dos poços adicionando-se 300uL de tripsina (2,5g/L - Atividade 1:250), incubando-se por 3min a 37 °C e adicionando-se 1700µL RPMI-1640 sem soro, homogeneizando e coletando em tubos de 15 mL. A solução de células foi centrifugada em baixa rotação (1200 rpm) e o meio foi removido. Em seguida foi adicionado 1 mL de PBS, solubilizando o *pellet* suave homogeneização para remoção de tripsina residual, novamente submetido a em baixa rotação em centrífuga e após a precipitação o PBS foi removido. O pellet resultante foi solubilizado em 1 mL de meio RPMI-1640 sem soro. A concentração celular foi confirmada por contagem em hemacitômetro.

Para o ensaio de migração, foi adicionado, primeiramente 500µL RPMI-1640 com soro no poço, em seguida o inserto (*Transwell*[®] de 6,5 mm, poros de 8,0µm- Cat#3458) no poço com meio e então foi adicionada 10^5 células em

300µL RPMI-1640 sem soro no inserto, com cuidado para não transbordar. O sistema então foi incubado por 24h a 37°C em estufa de CO₂.

Para o ensaio de invasão foi incubada placa e o inserto em RPMI-1640 sem soro *overnight* a 37°C, passado o período o meio foi removido. Foi adicionado, primeiramente 700µL RPMI-1640 com soro no poço, em seguida o inserto (Corning® Matrigel® 8 Micron - Cat#354480) no poço com meio e então foi adicionada 10⁵ células em 500µL RPMI-1640 sem soro no inserto, com cuidado para não transbordar. O sistema então foi incubado por 24h a 37°C em estufa de CO₂.

Encerrado o período de incubação para ambos foi feita a remoção do meio e foi feita fixação e coração, da seguinte forma: os insertos foram lavados com PBS, mergulhando-os num béquer com PBS, em seguida foram incubados por 2min PFA 4% na própria placa, lavados outra vez com PBS. Em seguida foram incubados por 20min em metanol PA na placa e novamente lavados com PBS. Por fim foram incubados, na placa, em solução de cristal violeta (0,5 g de pó de cristal violeta, 80 mL de água destilada, 20 mL de metanol) e lavados com PBS. Cada inserto então foi secado com algodão muito gentilmente e depois deixado a temperatura ambiente.

Os resultados foram visualizados em microscópio e registrado em câmera, programa *ImageJ* foi utilizado para análise de migração e invasão de células. O Cada experimento teve um n=3 em duplicatas técnicas, as análises estatísticas foram feitas utilizando-se o teste Mann-Whitney não pareado.

2.13. Ensaio xenográfico

Esse ensaio foi realizado em colaboração com a aluna de doutorado Thalita Bueno, orientada pela professora Daniela Bassères, e o Dr. Luís Bruno Cruz em nosso laboratório, orientado pelo professor Eduardo Reis.

A indução de tumores xenográficos subcutâneos foi realizada em 5 (cinco) animais, camundongos BALB/c *Nude* de 45 dias de idade imunossuprimidos e monitorados a cada 3 (três) dias durante 60 dias. Uma solução de 100µL de PBS com 10% de Matrigel® *growth factor reduced* (Corning) e contendo 1x10⁶ células AsPC-1 (transfectadas previamente com siCTRL ou siUCA1) foi preparada e administrada no subcutâneo (s.c.) dos flancos dos

animais (flanco direito com siRNA controle e flanco esquerdo com siRNA UCA-1) previamente anestesiados com cetamina (100mg/Kg) e xilazina (20mg/Kg) (Ceva). O dia no qual aconteceu a injeção foi considerado o dia 0 (zero). O volume tumoral foi monitorado e quantificado a partir da mensuração de duas dimensões com auxílio de um paquímetro manual. O volume sempre foi expresso em mm³ de acordo com a fórmula V (volume) = (comprimento x largura²) / 2, sendo que o comprimento foi considerado como sendo o maior das duas dimensões mensuradas no tumor (X. Yang et al. 2015).

2.14. Ensaio cometa alcalino

Este ensaio foi realizado em colaboração com a Dra. Lilian Cristina Russo Vieira, pós-doutora orientada pelo Prof. Fabio Luis Forti do Laboratório de Sinalização em Sistemas Biomoleculares.

Células AsPC-1 tratadas com siRNA e seus respectivos controles em duplicata na quantidade de 6×10^4 células foram semeadas em placas 12 poços, sendo que cada poço possui 3,8 cm² de área de crescimento celular, e incubadas por 48h. Então foram submetidas a irradiação não ionizante gama na dose de 10Gy. A exposição foi realizada em irradiador experimental de Cobalto-60, modelo Gamacell 220, localizado no Instituto de Pesquisas Energéticas e Nucleares da USP (IPEN). As condições do experimento foram as seguintes: amostras controle (não irradiadas) e amostras que aos 15min, 30min, 1h e 2h depois da irradiação tiveram as células removidas das placas por processo enzimático com tripsina formando assim uma suspensão de células, dessa 30 µL foram adicionados a 100 µL de solução de agarose 0,5% (m/v) de baixo ponto de fusão, homogeneizados e a mistura foi depositada em lâminas. As lâminas foram previamente preparadas com uma finada camada de agarose 1,5% (m/v). As suspensões de células em agarose foram então recobertas com lamínulas de microscopia, para assegurar um espalhamento homogêneo das células sobre a superfície da lâmina e incubadas *overnight* a 4 °C em solução de lise (2,5 M NaCl, 100 mM EDTA e 10 mM Tris, pH10) para rompimento de membranas. Em seguida foram submetidas à desnaturação do DNA (30 min, 4 °C) e subsequente eletroforese em tampão de pH alcalino (300 mM NaOH e 1 mM EDTA, pH>13) com 25 V e 300 mA por 30 min, a 4 °C. Após a eletroforese, as lâminas foram

mergulhadas em solução de neutralização (0,4 M Tris, pH 7,5) e incubadas por 15min seguido mergulhadas em etanol absoluto por 5min da fixação. Para a leitura, as lâminas foram submetidas a tratamento com EtBr (20 µg/mL) para visualização do DNA em gel. Cada lâmina foi visualizada e registrada em imagem proveniente de microscópio de fluorescência. As imagens de cada lâmina foram analisadas através do software *Komet6* (Andor), o qual afere a quantidade de dano ao DNA genômico e expressa sob a unidade de “momento de cauda” (*OliveTailMoment*), que é dado pela multiplicação da percentagem de DNA na cauda com o tamanho da cauda. Para cada amostra foram analisadas 100 células, sendo que foram 50 células observadas em cada réplica técnica (Magalhães et al. 2018).

3. Objetivos

3.1 Objetivo geral

Realizar uma análise com alta-resolução do transcriptoma no adenocarcinoma de pâncreas (PDAC) para identificar novos RNAs não codificadores e variantes de *splicing* e alterações transcricionais associados à malignidade

3.2 Objetivos específicos

- Gerar bibliotecas de RNA total para NGS (RNAseq) a partir de um conjunto de 14 amostras cirúrgicas pareadas de PDAC (T) e o respectivo tecido não tumoral adjacente (N).
- Validação de novos transcritos, novas isoformas e lncRNA aberrante em PDAC identificados após a análise computacional dos dados de RNAseq.
- Validação da expressão diferencial de novos transcritos e lncRNAs com expressão aberrante em PDAC em modelos de PDAC
- Avaliação funcional *in vitro* e *in vivo* de lncRNAs candidatos no fenótipo tumoral

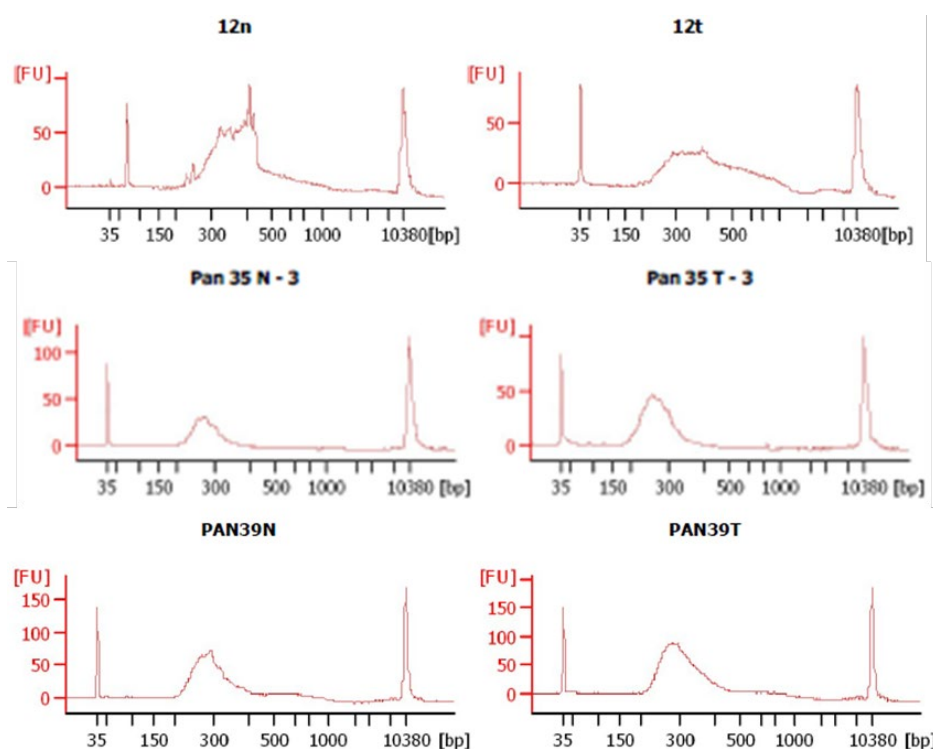
4. Resultados

4.1. A produção de bibliotecas de RNA-seq e sequenciamento preliminar

A produção de bibliotecas de RNAseq das 28 amostras cirúrgicas de pâncreas, sendo 14 amostras pareadas (Tabela 1) com as amostras tumorais (T) e as respectivas amostras não tumorais pancreáticas adjacentes (N) foi realizada com sucesso, no tamanho médio adequado (300bp) para o sequenciamento em plataforma Illumina (Figura 6). O perfil obtido por eletroforese capilar das 28 bibliotecas geradas está no Anexo1.

O sequenciamento preliminar em plataforma MiSeq, Illumina, produziu mais de 30 milhões de *reads* que passaram no filtro de qualidade do equipamento, com uma distribuição homogênea entre as amostras analisadas indicando boa precisão na quantificação das bibliotecas. A partir das bibliotecas de cada amostra foram montados 3 (três) lotes, os quais eram um “*pool*” de bibliotecas geradas partir de RNA total depletado de rRNAs.

Figura 6 – Perfil da eletroforese capilar das bibliotecas de cDNA de PDAC (T) e tecido não tumoral (N) geradas a partir de RNA total depletado de rRNAs. Pode-se observar o perfil de 3 (três) pares de amostras, cada par N e T é apresentado representativamente de cada um dos lotes de amostras sequenciadas.



4.2. Sequenciamento HiSeq

O sequenciamento em plataforma HiSeq 1500 gerou um total de 484.844.825 *reads*, com uma média de 17.182.995 de *reads* por amostra. Desses um total de 412.769.449 *reads* foram mapeados, com isso a média de mapeamento foi de 85%, uma cobertura média de 3,4GB do transcriptoma do PDAC, considerado satisfatório. Na tabela 7 pode-se observar a compilação desses dados para cada biblioteca (Sosa 2019).

Tabela 7 – Número total de *reads* gerados pelo sequenciamento para cada biblioteca; número total e porcentagem *reads* mapeados na referência.

Amostra	Nº total de <i>reads</i>	Nº total de <i>reads</i> mapeados	Porcentagem de <i>reads</i> mapeados
1N	25.465.297	22,383,996	88
1T	21.581.898	20,783,368	96
2N	16.819.523	9,553,489	57
2T	13.048.291	12,539,408	96
3N	13.506.679	12,074,971	89
3T	13.850.211	12,645,243	91
4N	22.893.874	20,077,927	88
4T	21.060.319	19,670,338	93
5N	12.668.831	9,032,877	71
5T	14.164.307	13,073,655	92
6N	13.639.049	8,565,323	63
6T	12.944.700	9,223,748	89
7N	11.636.296	10,728,665	92
7T	14.659.909	12,944,700	88
8N	17.334.453	10,626,020	61
8T	18.715.428	16,525,723	88

9N	11.485.387	10,061,199	88
9T	14.639.918	13,878,642	95
10N	21.180.027	18,045,383	85
10T	12.782.377	10,852,238	85
11N	21.739.600	20,348,266	94
11T	20.274.193	19,361,854	96
12N	17.555.622	7,777,141	44
12T	22.347.733	21,096,260	94
13N	15.073.484	12,857,682	85
13T	19.683.721	17,498,828	89
14N	20.070.073	16,999,352	85
14T	24.023.625	23,543,153	98
Média	17.182.995	14,705,530	85

4.3. Análises do transcriptoma detectou novos transcritos, novas formas de splicing e lncRNAs diferencialmente expressos no PDAC

A análise do RNAseq do transcriptoma das 28 bibliotecas de RNA total pareadas resultou

Tabela 8 – A reconstrução do transcriptoma pancreático. Número de loci e transcritos detectados nas diferentes etapas da montagem. Primeiro, os transcritos das amostras tumorais e não tumorais foram filtrados para remover RNAs curtos e de baixa expressão (“Filtro a”). A continuação as montagens foram combinadas para gerar o transcriptoma do pâncreas (“Mescla”) que foi filtrado para remover novas isoformas com exons aberrantes (n=8.080) e outros possíveis artefatos. Transcritos intergênicos e antisensos não anotados mono-exônicos (n=18.453 e 1.515, respectivamente) não foram considerados para análises subsequentes. O conjunto final (“Filtro b”) contém 41.341 transcritos conhecidos do GENCODE correspondentes com 17.834 *loci* (genes), assim como 6.710 transcritos não anotados (11,3% do total) de 4.359 *loci* (22,3% do total). A maioria dos transcritos não anotados (>90%) são variantes de splicing

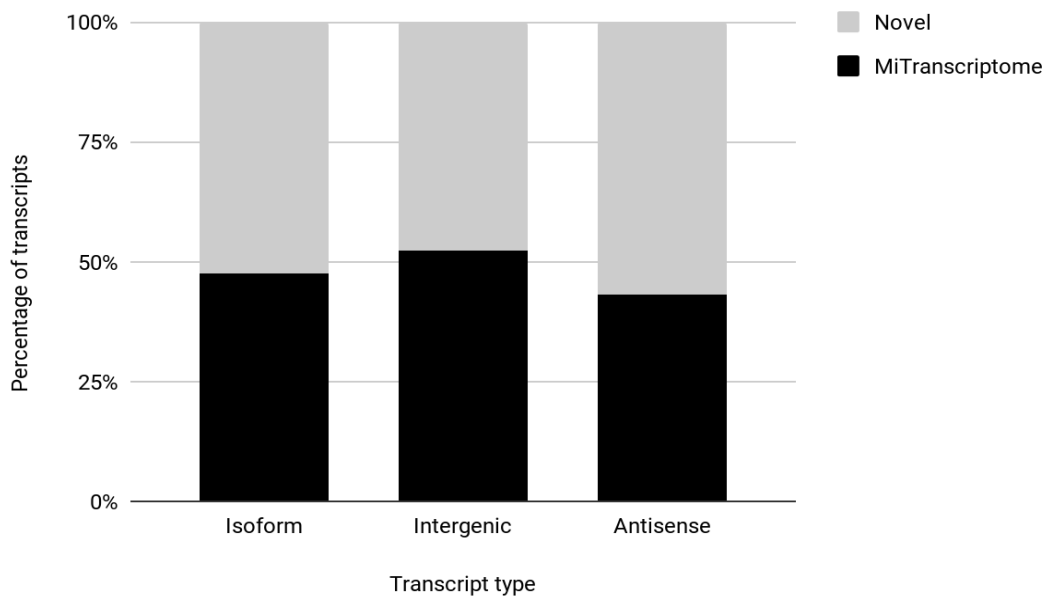
de genes conhecidos, uma proporção menor intergênicos (5%) ou antisensos (3%) em relação aos transcritos do GENCODE.

	<i>Loci</i>		<i>Trasncritos</i>	
	Não-tumoral	Tumor	Não-tumoral	Tumor
Total	66.307	92.761	149.586	198.830
Filtrados set (a)	28.081	30.625	58.762	73.613
Mescla	41.095		90.522	
Filtrados set (b)	21.668		62.474	
Anotados no GENCODE	17.834		41.341	
<i>Protein-coding</i>	13.560		30.771	
<i>Processed transcript</i>	3.951		5.993	
<i>Pseudogene</i>	1.540		1.578	
<i>lncRNA</i>	1.076		1.473	
<i>Antisense RNA</i>	1.196		1.526	
Não anotados	4.358		6.710	
<i>splicing variant</i>	3.935		6.187	
<i>intergenic RNA</i>	298		339	
<i>antisense RNA</i>	146		184	
<i>Other*</i>	8.770		14.423	

* A categoria "Other" inclui as seguintes classes de transcritos: *miscRNAs*, *3 prime overlapping ncRNAs*, *IG genes*, *Mt rRNA*, *non stop decay*, *retained intron*, *nonsense mediated decay*, *sense overlapping*, *snoRNA*, *snRNA*, *TR genes*, *possible pre-mRNA fragments*, *intronic RNAs*, *possible polymerase run-on fragments*.

Os resultados da montagem do transcriptoma do PDAC utilizando o catálogo do GENCODE foi comparado então ao catálogo do miTranscriptome (miT) (Iyer et al. 2015), para verificar a sobreposição com transcritos já detectados anteriormente na literatura e depositados no miT. O miT é um banco de dados no qual foram depositados dados de 7.256 bibliotecas de tecidos tumorais, normais e linhagens celulares, chegando numa reconstrução com 91.013 genes (sendo 68% deles lncRNAs). A nossa reconstrução teve uma sobreposição de aproximadamente 47% dos 6.187 potenciais novas isoformas de *splicing*, foram também detectados no miT 52% dos 339 potenciais novos transcritos intergênicos e 43% dos 184 novos antisense (Figura 9).

Figura 9 – Detecção dos transcritos não anotados no GENCODE detectados na montagem final dos nossos dados de RNAseq de PDAC com o catálogo miTranscriptome (miT).



Os transcritos reconstruídos que não foram anotados no GENCODE, considerados novos, foram comparados com a referência de um banco de dados independente, o miTranscriptome (miT). A sobreposição foi avaliada para as classes *Isoform* (*splicing variant* na tabela 8), *Intergenic* (intergenic RNA) e *Antisense* (antisense RNA). Foram detectados no miT: 47% dos 6.187 potenciais novas isoformas de *splicing* (*Isoform* no gráfico), foram também detectados no miT 52% dos 339 potenciais novos transcritos intergênicos (*Intergenic* no gráfico) e 43% dos 184 novos antisense (*Antisense*).

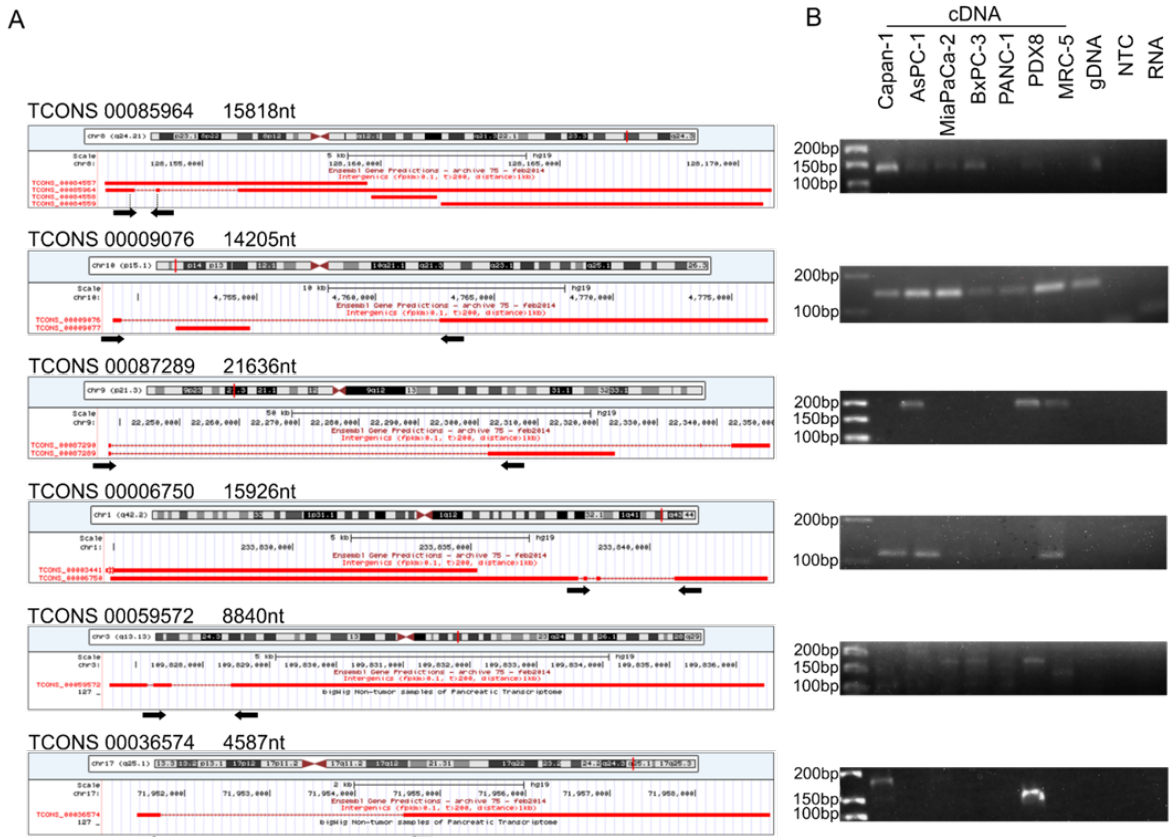
4.4. Novos transcritos intergênicos detectados em PDAC

A montagem reconstruiu um total de 339 transcritos intergênicos, desses foram selecionados aqueles superexpressos em PDAC em relação as respectivas amostras não tumorais ($FC > 1$) e cuja expressão fosse significativamente aumentada ($p\text{-valor} < 0,001$). Como a montagem desses transcritos não faz uso de anotações de genes existentes selecionamos transcritos multiexônicos, ou seja, com evidências de *splicing*. Esse filtro foi adotado para eliminar possíveis artefatos da montagem das análises, pois os transcritos monoexônicos podem ser RNAs não processados capturados na construção das bibliotecas de RNAseq. Com isso chegamos a um conjunto de 6 (seis) novos transcritos intergênicos (Figura 10A). Os novos transcritos intergênicos são identificados em nossa montagem como como TCONS00084995; TCONS00009076; TCONS00087289; TCONS00006750; TCONS00059572; e TCONS000036574, todos eles demonstraram serem não codificadores por análise de conservação e estrutural bem como não terem evidência de fase aberta de leitura (ORF - *open reading frame*) (TCONS00084995; TCONS00009076 e TCONS00059572) ou ter ORF mas sem evidência de codificar pequenos peptídeos (TCONS00087289; TCONS00006750 e TCONS00059572). Pela análise de expressão diferencial TCONS00084995 apresentou os seguintes parâmetros: $\log_2(FC) = 3,65$; TPM = 2,75 e $P_{adj} = 4,9 \times 10^{-13}$, o transcrito está anotado no miTranscriptome. TCONS00009076 apresentou os seguintes parâmetros: $\log_2(FC) = 6,88$; TPM = 26,10 e $P_{adj} = 2,13 \times 10^{-6}$, este transcrito é considerado uma nova forma de *splicing* desse locus anotado no miTranscriptome. O novo intergênico TCONS00087289 tem uma DE de $\log_2(FC) = 3,44$; TPM = 57,70 e $P_{adj} = 1,3 \times 10^{-3}$ e possui exatamente a mesma estrutura que foi reconstruída em nossa reconstrução sustentada pelo miTranscriptome. TCONS00006750 tem a DE de $\log_2(FC) = 3,95$ significativamente aumentada em PDAC ($P_{adj} = 7,4 \times 10^{-3}$), a expressão média desse transcrito é de uma expressão média de TPM = 70,75 e uma AS de um transcrito anotação no miT para esse *locus*. O TCONS00059572, tem uma DE de $\log_2(FC) = 4,99$ significativamente aumentada em PDAC ($P_{adj} = 9,4 \times 10^{-4}$), e uma expressão média de TPM = 14,39; também é identificado como uma AS pelo miT. Já o TCONS000036574

tem expressão significativamente ($P_{adj} = 5,89 \times 10^{-6}$) aumentada em PDAC na ordem de $\log_2(FC) = 5,70$; porém não é anotado no miT.

Para a validação da existência desses transcritos foi realizada com sucesso uma PCR *end point* em cDNA proveniente de linhagens de células de PDAC, linhagem derivada de PDX, MRC-5 e controle genômico (gDNA de Capan-1), controle sem DNA (NTC) e controle utilizando RNA de Capan-1 (Figura 10B). Os *primers* utilizados flanquearam as regiões de borda de éxons adjacentes.

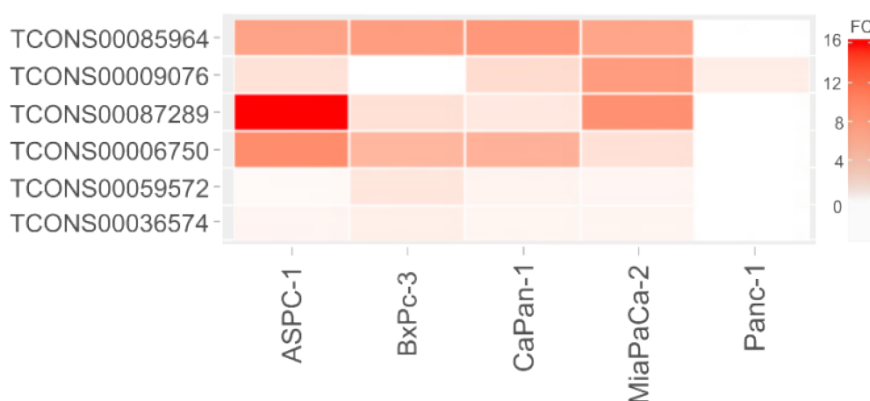
Figura 10 – Novos transcritos intergênicos selecionados e validados.



(A) tracks de visualização de cada um dos novos intergênicos no Genome Browser, as setas pretas indicam as regiões de borda de éxon flaqueadas por primers para validações por PCR *end point*. (B) Eletroforese em gel de agarose 2% da PCR *end point* a partir do cDNA de linhagens de PDAC (Capan-1, AsPC-1, MiaPaCa-2, BxPC-1 e PANC-1), linhagem derivada de PDX de PDAC (PDX8), linhagem de não tumoral imortalizada MRC-5 (Fibroblastos pulmonares), marcadores de tamanho molecular estão na *lane* 1 e controles (gDNA de Capan-1, NTC e no RT – RNA). As bandas obtidas tiveram os tamanhos esperados: TCONS00084995, 149bp; TCONS00009076, 145bp; TCONS00087289, 188bp; TCONS00006750, 114bp; TCONS00059572, 159bp; TCONS000036574, 153bp.

Os mesmos *primers* foram utilizados para uma mensuração da expressão relativa por RT-qPCR nas linhagens de PDAC (Figura 11). As linhagens utilizadas foram Capan-1, AsPC-1, MiaPaCa-2, BxPC-1 e PANC-1, o resultado demonstrou que a recapitulação da expressão dos 6 (seis) novos transcritos intergênicos observada no RNA-Seq não é observada em todas as linhagens, sendo que, cada transcrito não foi detectado em pelo menos 1 (uma) das linhagens. No caso do TCONS00059572 e TCONS00036574, os quais não foram detectados em PANC-1, quando expressos o foram em níveis muito baixos nas outras linhagens.

Figura 11 - *Heatmap* das expressões relativas dos novos transcritos intergênicos validados em linhagens celulares de PDAC por RT-qPCR.

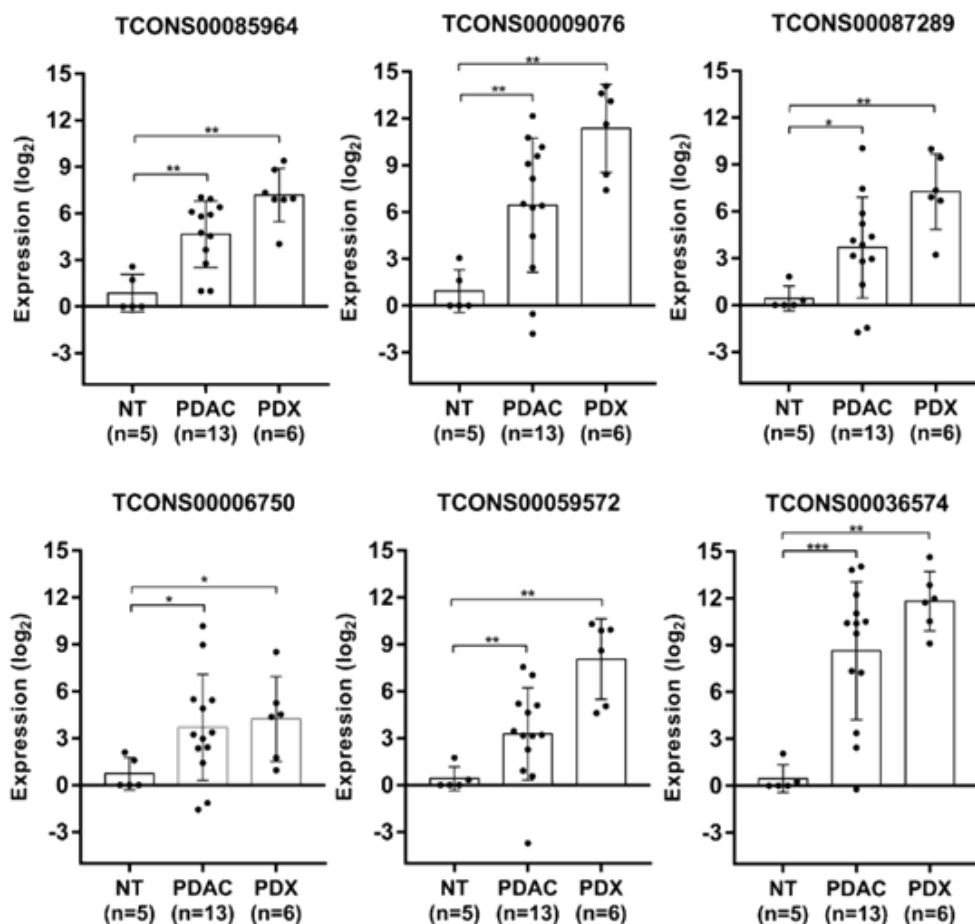


As expressões de cada transcrito foram medidas por RT-qPCR em triplicata e normalizadas em relação ao menor valor de expressão detectável, a expressão de TCONS00059572 em AsPC-1, essa normalização foi transformada em uma matriz contendo o *fold change* (FC) de cada expressão em relação ao normalizador e representada em *Heatmap*.

As expressões aberrantes constadas nos dados de RNAseq foram validadas por RT-qPCR em um conjunto de amostras cirúrgicas pancreáticas não tumorais (N), PDAC (T) e PDX (Figura 12). O conjunto de amostras N e T não foram pareados e continham alguma sobreposição com o conjunto de amostras utilizadas para produção do RNA-seq. Uma vez que as amostras T são heterogêneas (constituídas de RNA extraído de células do tumor e estroma) a expressão relativa foi também observada em um conjunto de PDX de PDAC, e pode ser confirmando que a expressão aberrante pode ser observada em células

tumorais do PDX e que esse modelo recapitulou níveis significativamente aumentados em relação a amostras N.

Figura 12 – Validação da expressão relativa aberrante por RT-qPCR dos novos transcritos intergênicos em amostras pancreáticas não-tumorais (NT), PDAC e PDX.

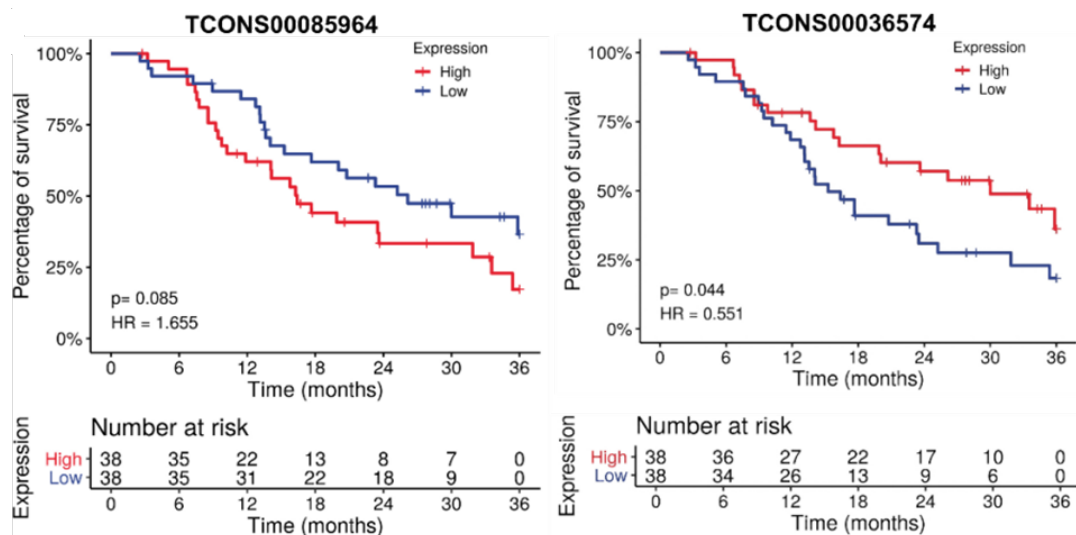


As expressões relativas de cada transcritos medida por RT-qPCR em um conjunto de amostras pancreáticas não tumorais (NT), amostras tumorais (PDAC) e em conjunto de PDX de PDAC. A expressão relativa foi calculada em relação a média das amostras N (* P-valor $\leq 0,05$; ** P-valor $\leq 0,01$; *** P-valor $\leq 0,001$).

Análises de sobrevida (Kaplan-Meier) para um período de 36 meses apontaram que os novos transcritos TCONS00085964 e TCONS00036574 tem seus níveis de expressão significativamente (Pvalor $\leq 0,1$) correlacionados com alterações da sobrevida (Figura 13). O grupo que apresenta níveis de expressão de TCONS00085964 maiores que a mediana estão correlacionados com uma menor sobrevida quando comparado ao grupo cuja expressão foi menor que a mediana. Por outro lado, para TCONS00036574 a correlação foi inversa, o grupo

com níveis de expressão desse transcrito acima da mediana apresentou uma maior sobrevida quando comparado ao outro grupo.

Figura 13 – Análise da sobrevida para TCONS00085964 e TCONS00036574



A expressão de novos transcritos intergênicos foi significativamente (TCONS00036574, HR: 0,55 p = 0,044) ou marginalmente (TCONS00085964, HR: 1,66, p = 0,085) associada com alterações na sobrevida do paciente no conjunto de dados do ICGC (n=76).

4.5. Validação da expressão aberrante das novas formas de *splicing* (AS) em PDAC

A nossa reconstrução gerou 6.187 potenciais novas isoformas de *splicing* (AS) pelo catálogo do GENCODE, esse conjunto teve sobreposição de aproximadamente 47% com catálogo do miT. Os eventos de AS encontrados foram classificados em: 1.083 sítios alternativos 3' (A3 16,6%), 1.006 sítios alternativos 5' (A5 15,4%), 538 primeiro éxon alternativo (AF 8,3%), 87 último éxon alternativo (AL 1,3%), 62 éxons mutuamente excludentes (MX 1,0%), 1.831 retenções de íntrons (RI 28,1%) e 1.914 éxons ignorados (SE 29,4%). Outros 1.975 que não foram anotados pelo programa devido a sua complexidade e foram classificados como "Others" (Sosa 2019).

Como a montagem não fez uso de anotações de genes existentes, foram selecionados potenciais AS significativamente superexpressos em PDAC (Padj<0,001; log₂FC>2) em relação as respectivas amostras não tumorais e com

alguma evidência de alteração da ORF e cuja forma canônica do transcrito tivesse algum registro em contexto de câncer. Foram selecionados potenciais AS para os genes *MMP14* (TCONS00027381), *CAPN8* (TCONS00006585), *LIF* (TCONS00057528) e *OCT3* (TCONS00074972), os quais eram os transcritos com maior abundância dentre os selecionados pelo filtro de expressão. As existências dos AS selecionados foram validadas com sucesso por PCR *end point* (Figura 14).

O transcrito TCONS00027381 é considerado uma nova AS de *MMP14*, mais especificamente uma AS do transcrito canônico ENST00000311852.6 pelo catálogo do GENCODE, porém está contido no catálogo do miT (código de identificação T101069). Pela análise de expressão diferencial apresentou os seguintes parâmetros: $\log_2FC=2,46$; $P_{adj}= 1,24 \times 10^{-7}$; TPM = 1107,34. Foram detectados 4 (quatro) transcritos oriundos do *MMP14* no transcriptoma de PDAC, o candidato selecionado fora o mais expresso (TPM = 1107,34) os outros 3 (três) eram tiveram níveis muito baixos de expressão (TPM = 10,40; 0,67; e 3,425).

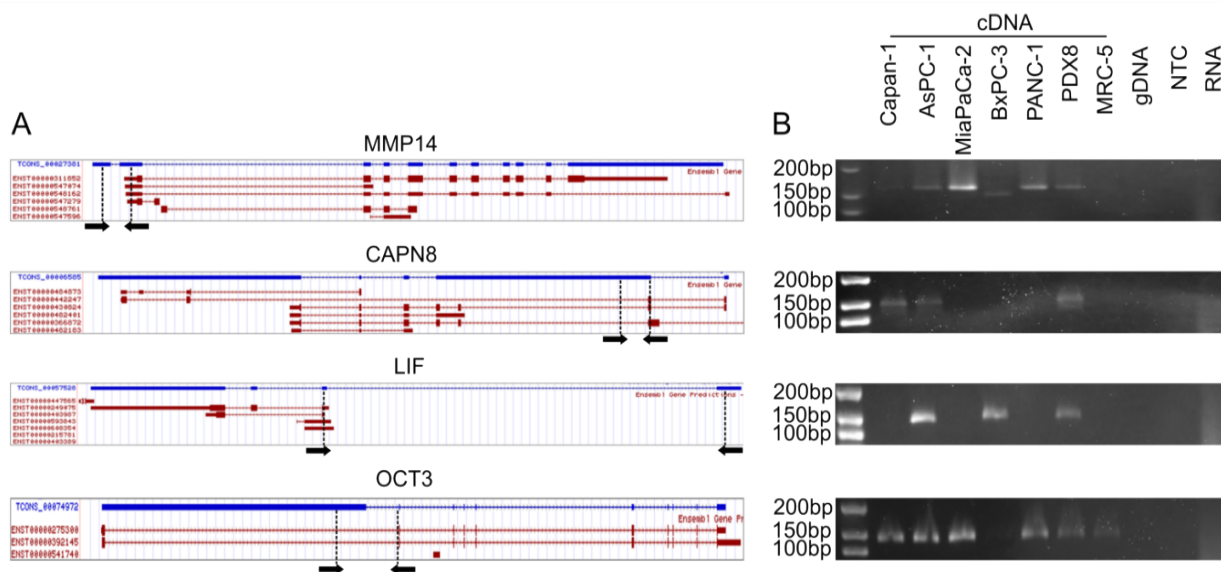
O transcrito TCONS00006585 foi classificado como uma nova AS de *CAPN8*, uma variação AS do transcrito canônico ENST00000366872.5, pelo catálogo do GENCODE, esse por sua vez apresenta uma sobreposição perfeita com uma referência do miT (T031330). Foi identificado como sendo significativamente superexpresso em PDAC ($\log_2FC= 5,34$; $P_{adj}= 5,11 \times 10^{-11}$; TPM = 516,15). Foram detectados 6 (seis) forma de *splicing* de *CAPN8* no transcriptoma de PDAC, sendo que o transcrito selecionado para validação é o *CAPN8* mais abundante (outros transcritos *CAPN8* TPM = 137,06; 56,79; 26,10; 1,07; 12,68).

O transcrito TCONS00057528, foi identificado como uma nova AS do gene *LIF*, uma variação do transcrito ENST00000249075.3, pelo catálogo do GENCODE, essa AS é sustentada pelo catálogo do miT (T232019), também significativamente superexpresso em PDAC ($\log_2FC= 2,16$; $P_{adj}= 9,3 \times 10^{-4}$; TPM = 164,78). Foram identificadas em nossa reconstrução 2 (duas) formas de *splicing* do gene *LIF*, a forma selecionada para validação tratou-se da mais expressa (outro transcrito tem um TPM = 7,79).

O transcrito TCONS00074972, é uma nova AS do gene *SLC22A3*, que tem como alias comum *OCT3*, trata-se de uma variação do transcrito canônico ENST00000392145.1. Essa nova forma de *splicing* é suportada pelo catálogo do

miT (T314629). Foi identificado como sendo significativamente superexpresso em PDAC ($\log_2FC = 2,88$; $Padj = 3,0 \times 10^{-4}$; $TPM = 192,43$). Foram detectados 4 (quatro) transcritos OCT3 no transcriptoma de PDAC, o candidato selecionado fora o mais expresso ($TPM = 192,43$) os outros 3 (três) eram tiveram níveis muito baixos de expressão ($TPM = 22,47$; $11,33$; NA).

Figura 14 – validação da existência das novas formas de *splicing* (AS)



(A) Tracks com a visualização das novas formas de splicing, as setas pretas indicam as regiões de borda de éxon flanqueadas por primers para validações por PCR end point. (B) Eletroforese em gel de agarose 2% da PCR end point a partir do cDNA de linhagens de PDAC (Capan-1, AsPC-1, MiaPaCa-2, BxPC-1 e PANC-1), linhagem derivada de PDX de PDAC (PDX8), linhagem de não tumoral MRC-5 e controles. As bandas tiveram os tamanhos esperados: MMP14, 200bp; CAPN8, 189bp; LIF, 175bp; OCT3, 175bp.

4.6. Validação da expressão aberrante de lncRNAs em PDAC

Para investigar quais lncRNAs tem expressão aumentada nas amostras de PDAC foram feitas uma série de análises dos dados de DE do RNAseq. A reconstrução chegou a 41.341 transcritos conhecidos (anotados no catálogo do GENCODE). As anotações do GENCODE atribuem a cada transcrito um biótipo, desses destacam-se na nossa reconstrução: *Protein-coding*, *Processed*

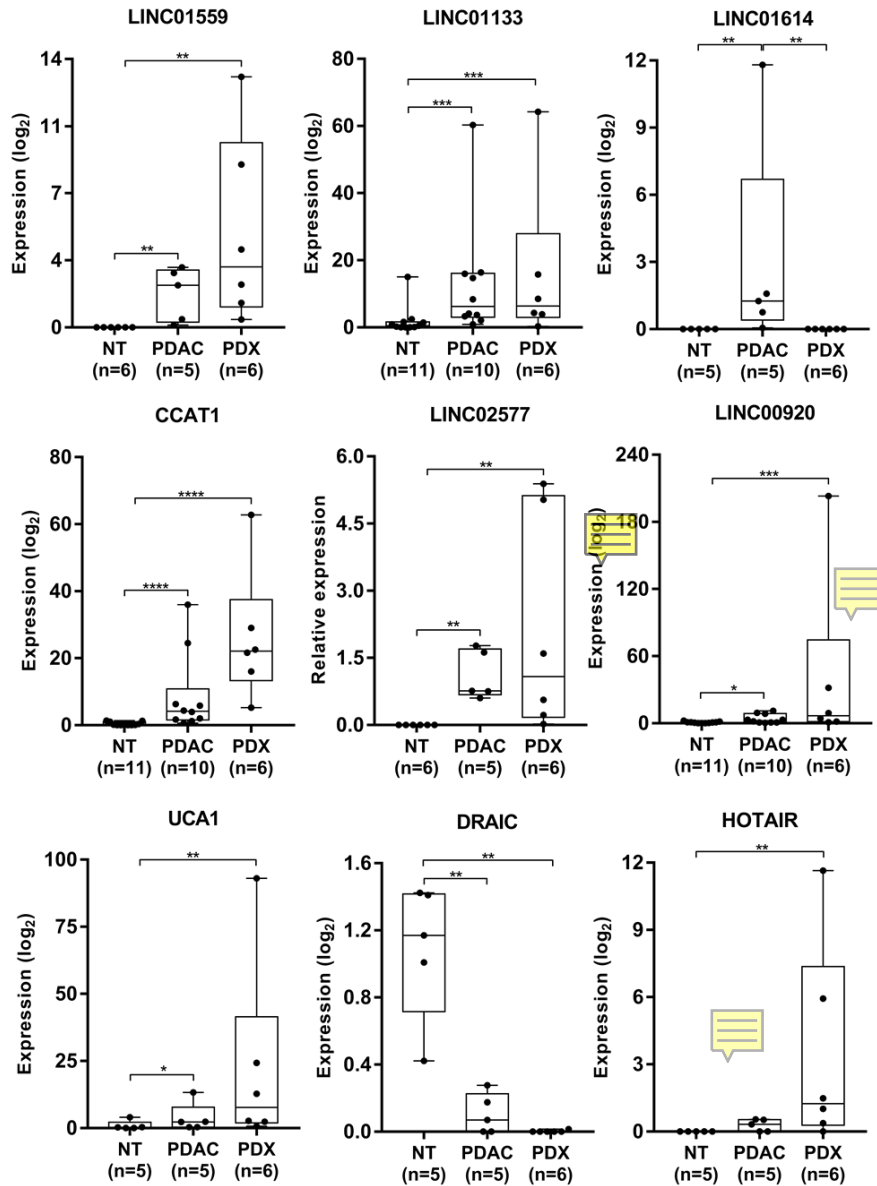
transcript, *Pseudogene*, *lncRNA*, *Antisense RNA* (Tabela 8, item 4.3). Outros biótipos foram agrupados como *other* (*miscRNAs*, *3 prime overlapping ncRNAs*, *IG genes*, *Mt rRNA*, *non stop decay*, *retained intron*, *nonsense mediated decay*, *sense overlapping*, *snoRNA*, *snRNA*, *TR genes*, *possible pre-mRNA fragments*, *intronic RNAs*, *possible polymerase run-on fragments*). Para estudarmos os lncRNAs foram selecionados para análise os biótipos *lncRNA* (n=1.473) e *Processed transcript* (n=5.993), pois o GENCODE define como *Processed transcript* os transcritos longos que não contém evidência de ORF. As mesmas análises foram feitas nível de gene. 17.834 genes reconstruídos estão anotados no GENCODE, desses 3.951 são classificados como *Processed transcript* e 1.076 são classificados como *lncRNA*. Com isso estabeleceu-se um conjunto de 5.027 genes de lncRNAs para as análises seguintes.

Ao conjunto de genes de lncRNAs foram aplicados filtros de expressão e significância estatística para selecionar apenas os superexpressos em PDAC. Considerando apenas os que tiveram um $P_{adj} < 0,001$ e $\text{Log}_2\text{FC} > 3,32$ estabeleceu-se o grupo de 532 genes *up-regulated* em PDAC. Desse conjunto *up-regulated* foram selecionados para validação em amostras N e T clínicas e PDX os 7 genes mais abundantes: *LINC01559*; *LINC01133*, *LINC01614*; *CCAT1*; *LINC02577*; *LINC00920* e *UCA1*, como controle experimental os genes *DRAIC* e *HOTAIR* foram incluídos no conjunto de validação. As expressões aberrantes dos lncRNAs foram confirmadas por RT-qPCR, *LINC01559*; *LINC01133*; *LINC01614*; *CCAT1*; *LINC02577*; *LINC00920* e *UCA1* demonstraram ser significativamente superexpressos nas amostras de PDAC (T) em relação as amostras não tumorais (N). Esses lncRNAs demonstraram ser mais enriquecidos nas amostras de PDX do que nas tumorais, exceto o *LINC01133*, que demonstrou uma média de *fold change* semelhante para amostras T e PDX e o *LINC01614* que não foi detectado nas amostras de PDX.

Os controles tiveram comportamento aberrante descrito em literatura; *DRAIC* já descrito com expressão diminuída em câncer de próstata (Rathinasamy and Velmurugan 2018), já era apontado como *downregulated* em PDAC e outros tumores gastrointestinais por análises *in silico* (Chakraborty 2014), *DRAIC* demonstrou ter a expressão diminuída em PDAC e não detectada em PDX por RT-qPCR. *HOTAIR* é descrito como superexpresso em PDAC (S.-Z. Yang et al. 2017) e não foi observada uma superexpressão significativa nas

amostras tumorais, porém significativamente *upregulated* em PDX (Figura 14). Esse resultado valida a DE observada por RNAseq dos lncRNAs selecionados.

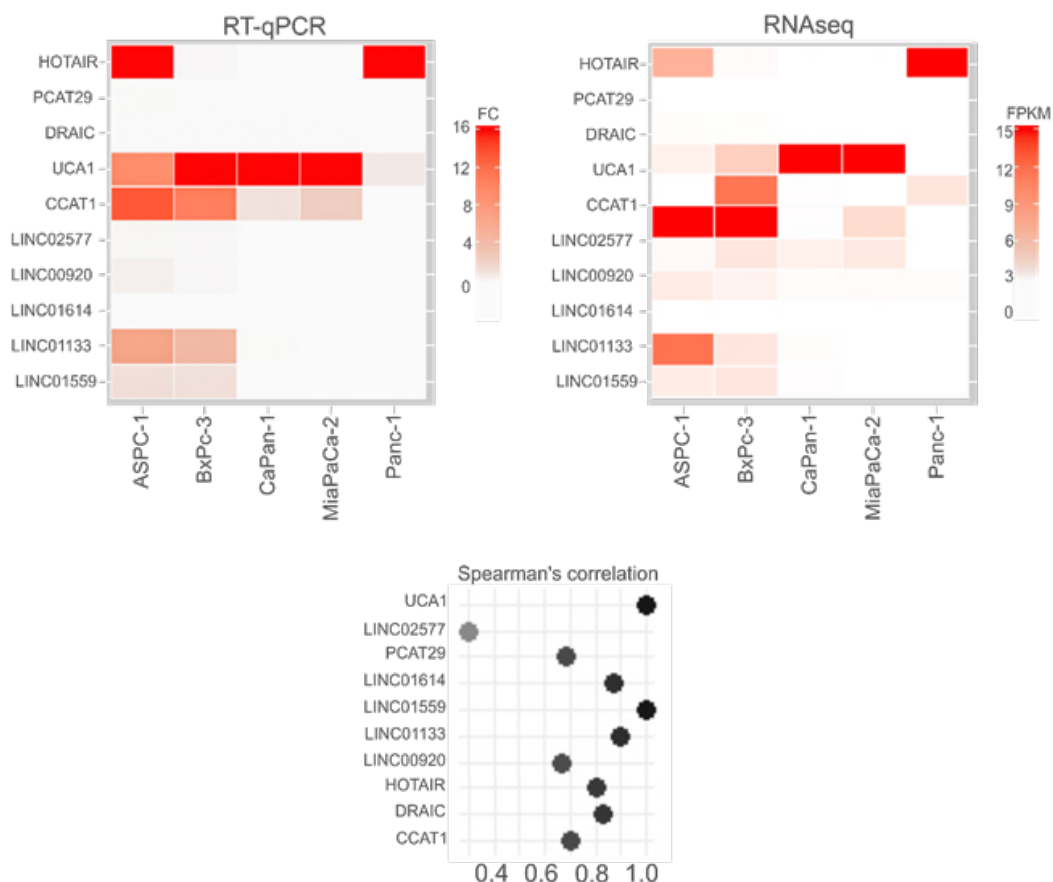
Figura 15 – A expressão relativa determinada por RT-qPCR dos lncRNAs confirma a expressão aberrante em tumor de PDAC dos lncRNA observada pelo RNAseq.



RNA proveniente de um conjunto diferente de amostras não tumorais (NT), tumorais (PDAC) e xenotumorais (PDX). LINC01559; LINC01133, LINC01614; CCAT1; LINC02577; LINC00920 e UCA1 são superexpressos em tumor quando comparados com amostras de tecido não-tumoral. Um conjunto de PDX de PDAC confirma e ratifica a expressão aberrante e determina se o lncRNA tem origem em células tumorais ou potencialmente do estroma tumoral, como LINC01614, cuja expressão não foi detectada em PDX. (* P-valor ≤ 0,05; ** P-valor ≤ 0,01; *** P-valor ≤ 0.001; ****P ≤ 0.0001).

Em seguida, foram verificadas as expressões dos lncRNAs candidatos por RT-qPCR em linhagens de PDAC; AsPC-1, BxPC-1, Capan-1, MiaPaCa-2 e PANC-1 para a seleção das linhagens celulares a serem utilizadas em estudos posteriores. Uma vez que todos os pares de *primers* utilizados possuem níveis de eficiência semelhantes as expressões de cada lncRNA para cada linhagem foi normalizado para expressão de LINC00920 em BxPC-3, o resultado da expressão relativa de cada lncRNA foi representado por *Heatmap*, além disso expressões absolutas dos mesmos lncRNAs, determinadas por RNA-seq, foram obtidas do banco de dados E-MTAB-2706 e representadas por *Heatmap*, ambas as expressões, por RT-qPCR e por RNAseq, demonstraram ter boa correlação (Figura 16). O gene PCAT29 foi incluído nessas análises como controle experimental de um gene *downregulated*. De todo o conjunto o LINC01614, assim como não foi detectado em PDX, não foi detectado tanto por RT-qPCR quanto por RNAseq as linhagens celulares de PDAC. Concluiu-se que AsPC-1 expressa relativamente os genes de interesse, sendo um modelo experimental adequado para estudos.

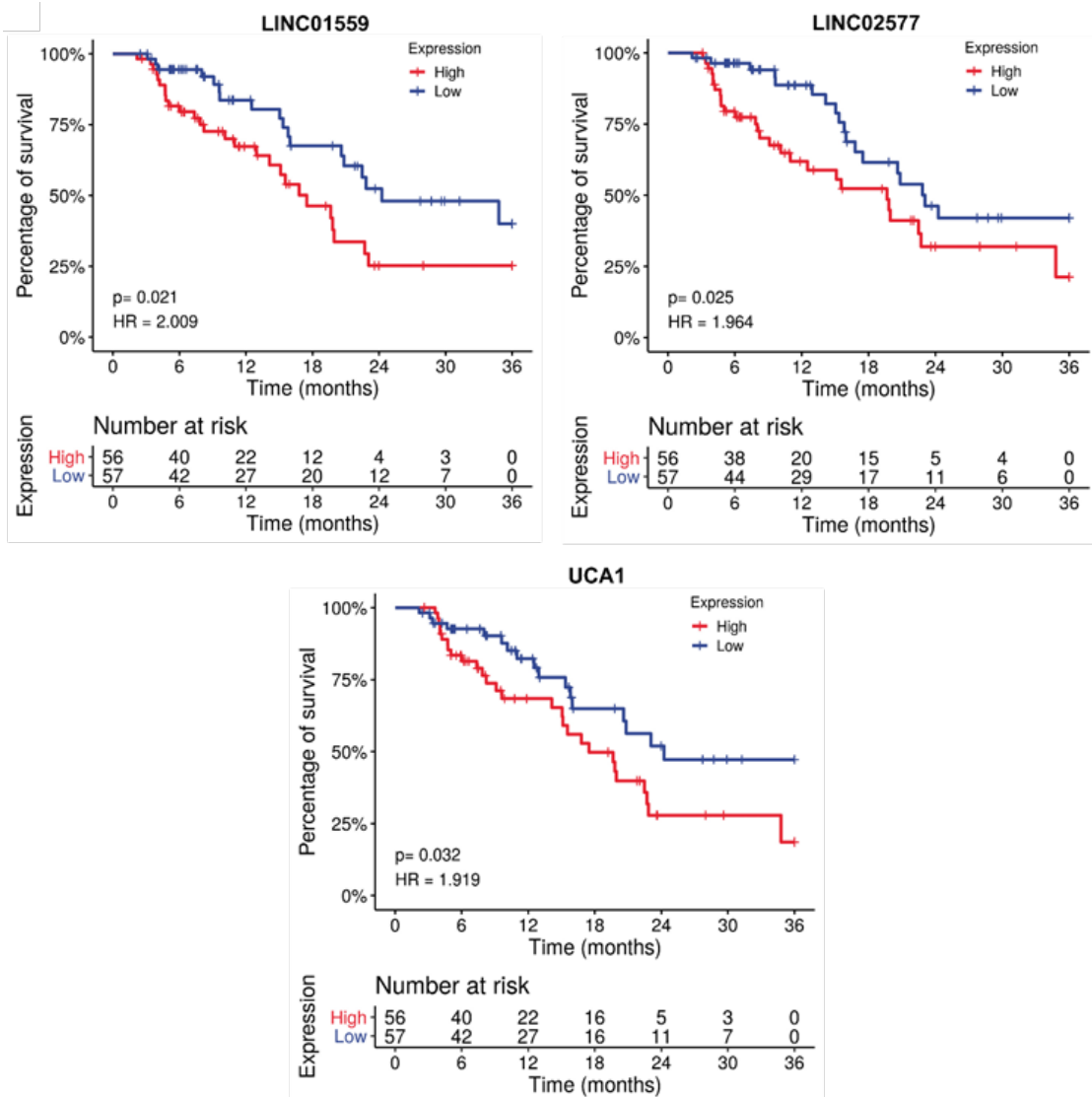
Figura 16 - *Heatmap* das expressões dos lncRNAs em linhagens celulares de PDAC.



As expressões relativas dos lncRNAs determinadas por RT-qPCR e normalizadas em função da expressão de DRAIC em AsPC-1. As expressões relativas foram medidas em triplicata, a expressão de cada lncRNA em cada linhagem foi normalizada e transformada em uma matriz contendo o *fold change* (FC) da expressão de cada lncRNA e representada em *Heatmap*. As expressões absolutas medidas em FPKM foram obtidas do banco de dados de RNA-Seq (E-MTAB-2706, NAR 2016) e foram representadas em *Heatmap*. A correlação (Spearman) da expressão medida por RT-qPCR e RNAseq foi medida para cada lncRNA.

Do conjunto selecionado (*LINC01559*; *LINC01133*, *LINC01614*; *CCAT1*; *LINC02577*; *LINC00920* e *UCA1*), foi constatado por análise de sobrevivência que as expressões de *LINC01559*, *LINC02577* e *UCA1* correlacionam-se com sobrevivência (Figura 17). O grupo que apresentou níveis de *LINC01559*, *LINC02577* e *UCA1* acima da mediana apresentou pior prognóstico.

Figura 17 – Análise de sobrevida para LINC01559, LINC02577 e UCA1.



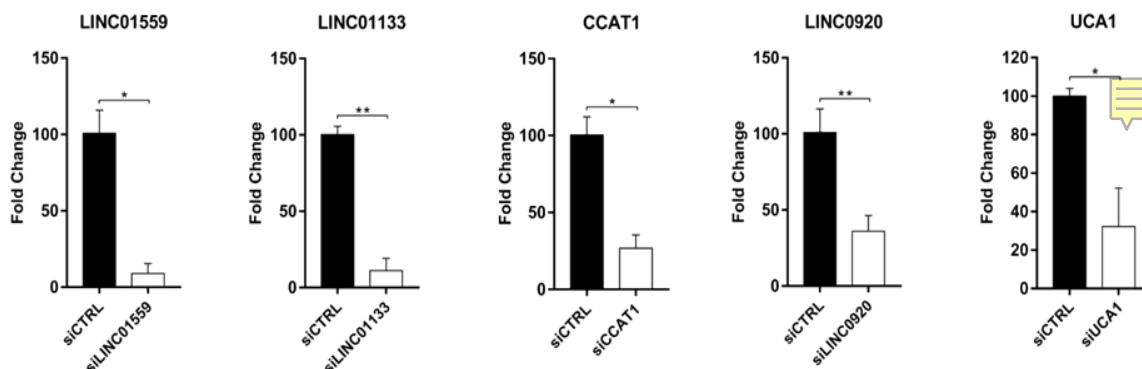
A partir dos dados do TCGA (PAAD-US) com $n=113$ pacientes foi medida a correlação por dos níveis de expressão de LINC01559, LINC02577 e UCA1 com alterações da sobrevida em pacientes em meses.

4.7. lncRNAs envolvidos com migração de invasão

Uma vez confirmada a expressão aberrante dos lncRNAs selecionados e validados (item 4.6) foram feitos ensaios para testar o envolvimento da expressão desses lncRNAs no fenótipo tumoral de migração e invasão. O conjunto de lncRNAs teve então a migração de invasão avaliadas em AsPC-1 com silenciamento transiente por siRNA para LINC01559; LINC01133; CCAT1;

LINC00920 e UCA1. Cada lncRNA foi silenciado com eficiência de 72 a 91% em AsPC-1 (Figura 18).

Figura 18 – *Knockdown* dos lncRNAs em linhagem AsPC-1

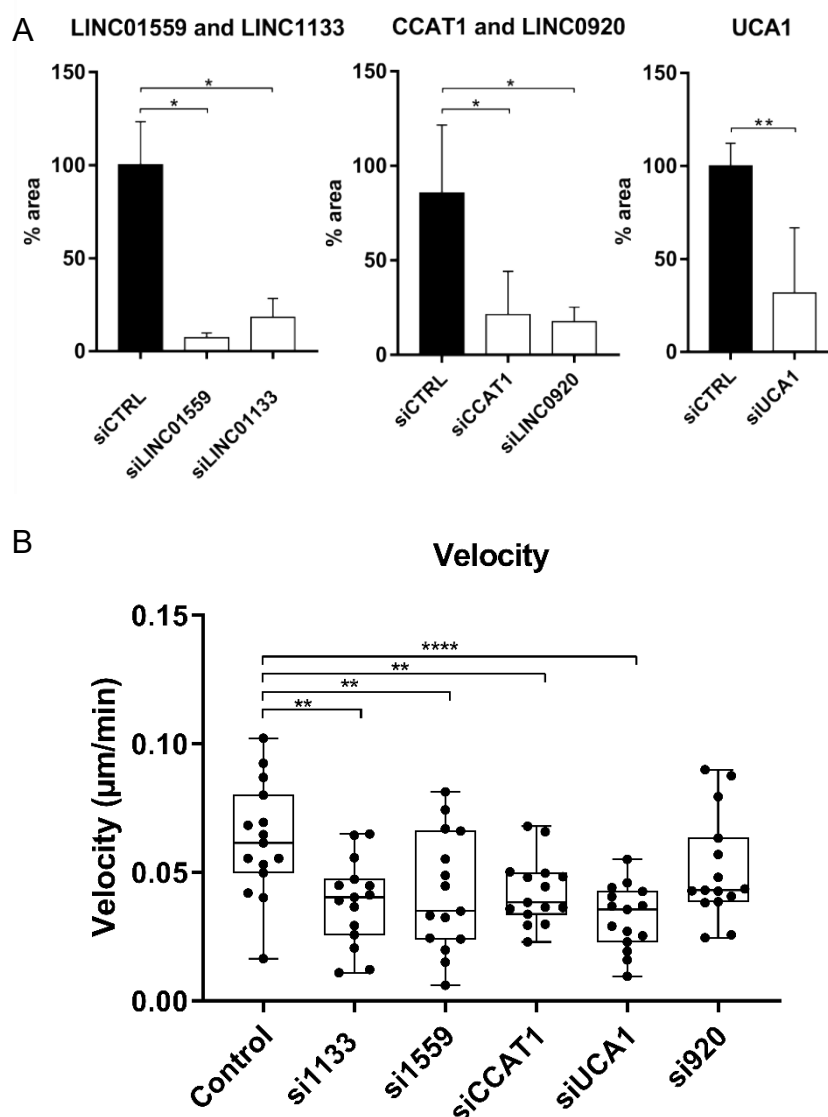


As silenciamento transiente por siRNA de linhagens de AsPC-1 dos lncRNAs com expressão aberrante em PDAC após 96h da transfecção. Foi observado o silenciamento de 91,05% para LINC01559; 72,82% para LINC01133; 73,21% para CCAT1; 64,07% para LINC0920 e 77,66% para UCA1 (* P-valor $\leq 0,05$; ** P-valor $\leq 0,01$).

Todos os lncRNAs demonstraram estar envolvidos com a capacidade de migração (Figura 19) e invasão (Figura 20), bem como a capacidade de proliferação (Figura 21) diminuídas comparadas aos controles transfectados com siCTRL. Foram observadas faixas de: 68 a 92% de redução das taxas de migração; 37 a 83% de redução das taxas invasão e 23 a 65% de redução das taxas de proliferação no contexto de silenciamento dos lncRNAs. O LINC00920 não demonstrou diferenças significativas no ensaio de proliferação, por isso seu resultado não foi incluído na figura 21.

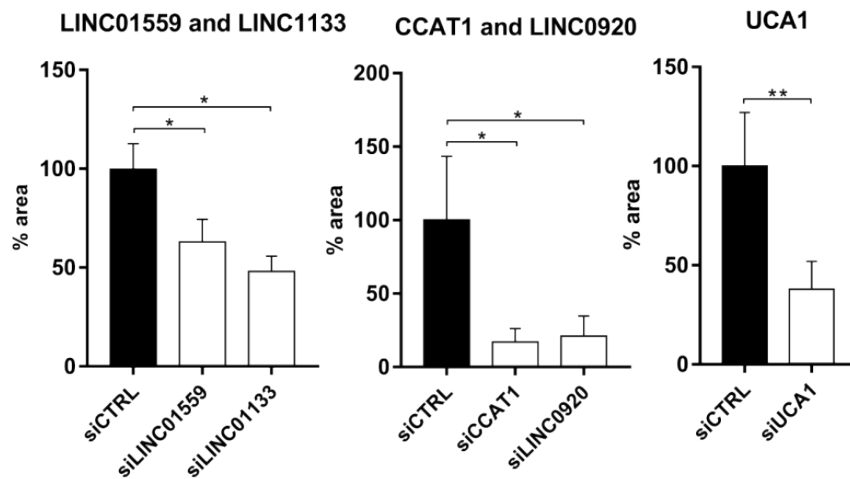


Figura 19 – Ensaio de migração *transwell* e imageamento.



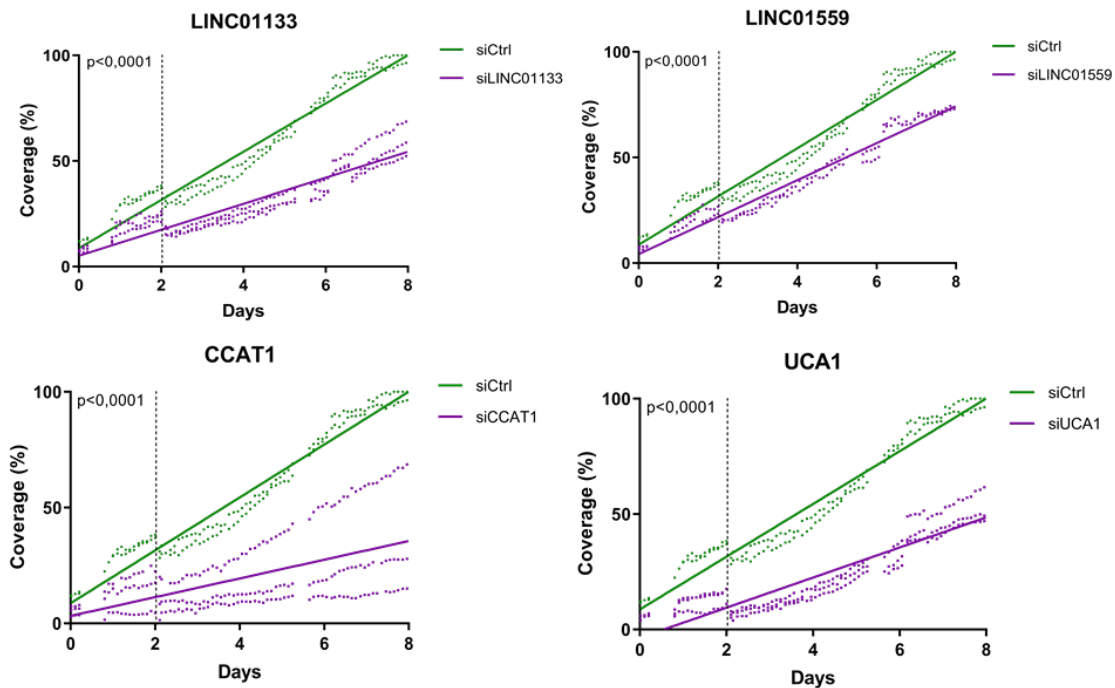
A capacidade de migração demonstrou ser significativamente reduzida em contexto de *Knockdown* de LINC01559; LINC01133; CCAT1; LINC00920 e UCA1. (A) Foi observada uma redução da capacidade de migração em ensaio *transwell* de 92,67% para LINC01559; 81,75% para LINC01133; 78,75% para CCAT1; 82,5% para LINC0920 e 68,29% para UCA1. (B) Avaliação da capacidade de migração pela análise das velocidades por um período de 6h para 15 células avaliadas em cada réplica e determinação das respectivas velocidades médias. (* P-valor \leq 0,05; ** P-valor \leq 0,01)

Figura 20 – Ensaio de invasão, *transwell*.



A capacidade de invasão demonstrou ser significativamente altera em contexto de *Knockdown* de LINC01559; LINC01133; CCAT1; LINC00920 e UCA1. Foi observada uma redução da capacidade de invasão de 37% para LINC01559; 52% para LINC01133; 83% para CCAT1; 79% para LINC0920 e 62% para UCA1 (* P-valor $\leq 0,05$; ** P-valor $\leq 0,01$).

Figura 21 – Avaliação da proliferação tumoral.



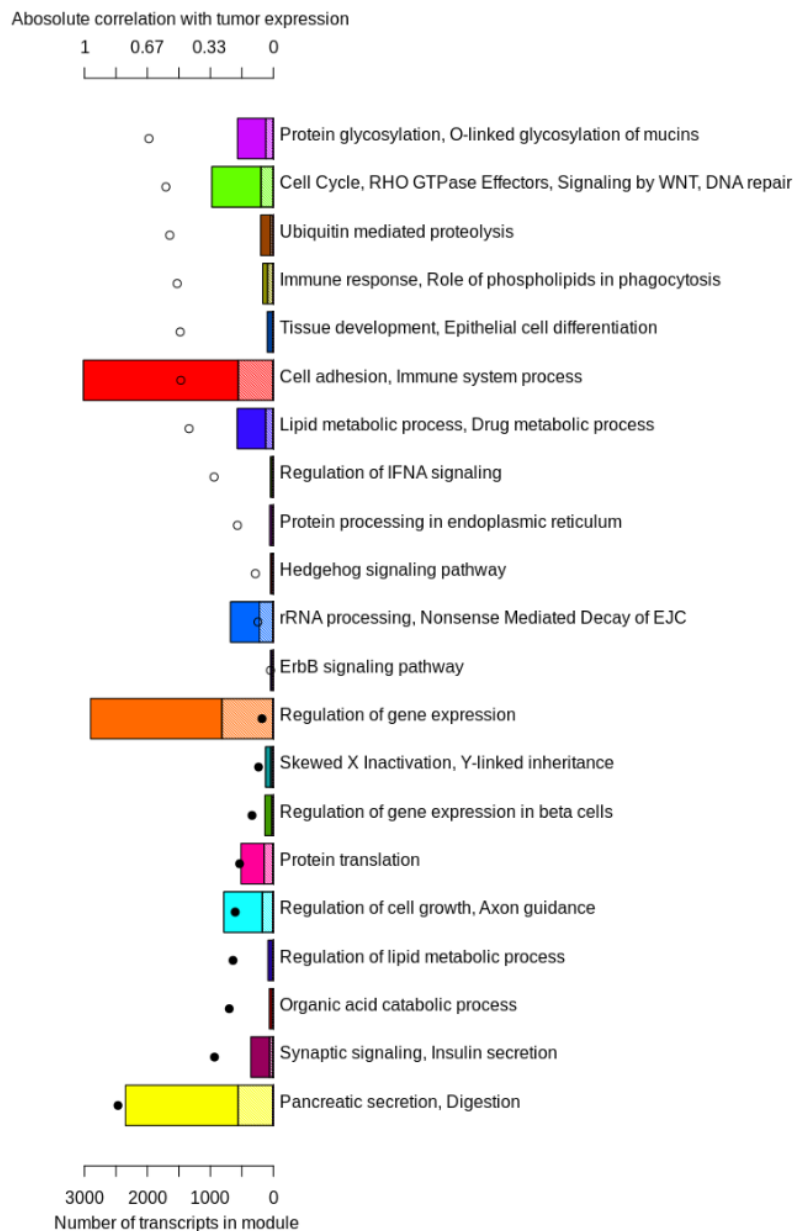
Ensaio de proliferação foi feito por método de imageamento ao vivo de fotos a cada 20min durante 8 dias. A reta roxa foi feita a partir da regressão linear da proliferação medida para um controle (n =2) de 2×10^4 células transfectadas com siCTRL. As retas verdes são a regressão linear da proliferação medida para um controle (n =3) em 2×10^4 células transfectadas com os siLINC01559, siLINC01133, siCCAT1 e siUCA1. Foram constatadas reduções das taxas de proliferação de 23% para siLINC01559, 46% para

siLINC01133, 65% para siCCAT1 e 43% para siUCA1. A significância desses valores foram verificadas utilizando análise de covariância (ANCOVA), ($P < 10^{-4}$) que compara as inclinações das retas.

4.8. Enriquecimento e WCGNA

Os lncRNAs são envolvidos em várias funções biológicas diferentes, por isso para se estudar os mecanismos em que eles estão implicados faz-se uma análise *guilt by association*. Essa análise busca inferir o possível papel dos lncRNAs correlacionando-os com mRNAs por uma rede de co-expressão. Posteriormente a análise verifica quais são as vias enriquecidas naqueles mRNAs co-expressos com o lncRNA em questão. As análises pelo WCGNA geraram 22 módulos de co-expressão (Figura 22), desse conjunto 2 (dois) módulos (Figura 23 e 24) que agrupavam candidatos foram avaliados com maior profundidade pelas características biológicas e por conterem transcritos selecionados nas outras análises do transcriptoma.

Figura 22 – Módulos de co-expressão e vias em que eles são enriquecidos.



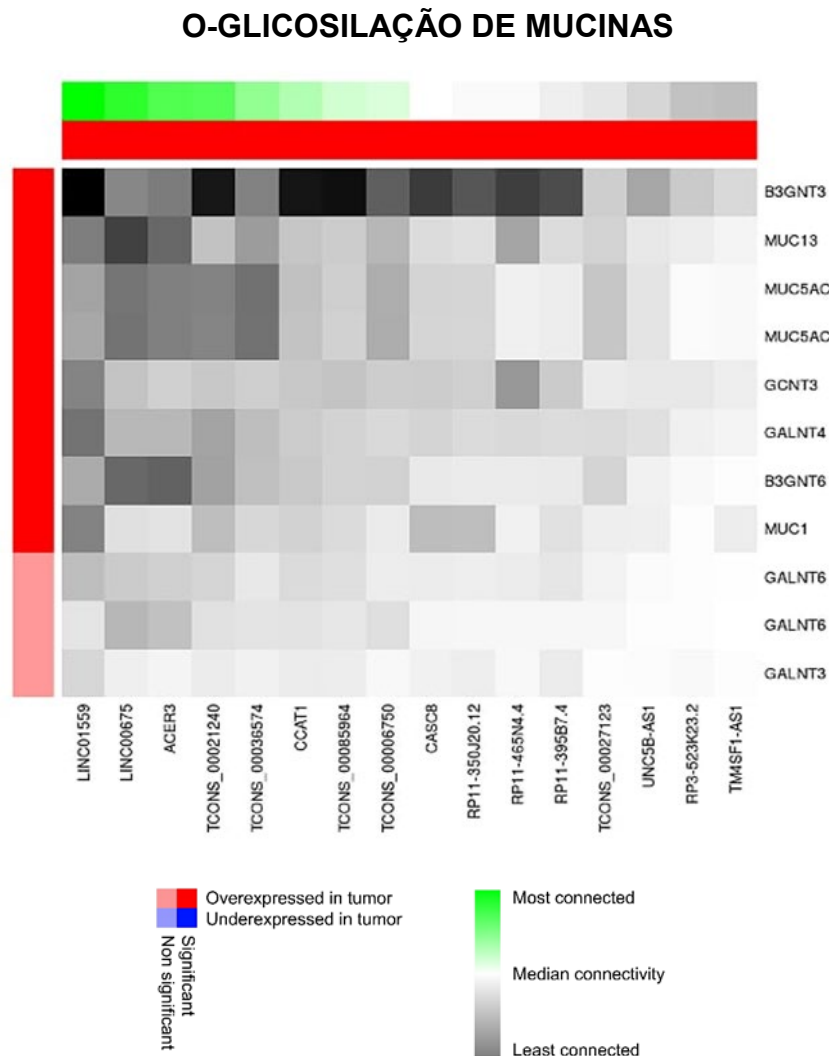
O comprimento das barras representa o número de genes contidos em cada módulo, com uma linha vertical indicando quantos são os não codificadores (eixo inferior), e para cada módulo é indicado por um círculo a média das correlações (Spearman) entre os genes e o fenótipo, os círculos vazios representam valores de correlação positivos os círculos pretos representam valores negativos (eixo superior).

A análise foi satisfatória. O *cluster* que recebeu a cor roxa demonstrou por esta análise independente estar mais bem correlacionado com o fenótipo tumoral, e quando analisado constatou-se estar entre as vias enriquecidas apresenta vias relacionadas à síntese de mucinas, classicamente relevantes no PDAC (Kaur et al 2013). Uma outra confirmação da qualidade da análise se dá pelo *cluster* amarelo, o qual é o mais negativamente correlacionado com o

fenótipo, este *cluster* apresenta a *Secreção Pancreática* entre as vias enriquecidas, tal correlação altamente negativa demonstra a perda de função do tecido tumoral quando comparado ao tecido não tumoral adjacente.

O LINC01614, cuja expressão não foi detectada em PDX e em linhagens celulares de PDAC, foi localizado no *cluster* vermelho de adesão celular. Os LINC01559, CCAT1 e os novos intergênicos TCONS00036574, TCONS00085964 e TCONS00006750, todos DE foram localizados no *cluster* roxo, o mais correlacionado com fenótipo tumoral. Esse *cluster* roxo é enriquecido para via de O-glicosilação de mucinas (Figura 23).

Figura 23 – *Heatmap* do módulo de coexpressão enriquecidos para a via de O-glicosilação de mucinas (*O-linked glycosylation of mucins*) com os lncRNAs DE do módulo.



As cores representam a distância na rede entre pares de genes, sendo sempre um lncRNA e um membro da via de “O-glicosilação de mucinas” (coding-RNA nas linhas e ncRNAs nas colunas, sendo pretos os mais próximos e brancos os mais distantes).

Todos os não codificadores mostrados são diferencialmente expressos, barras laterais indicam em azul os que estão mais expressos em amostras controle e em vermelho os mais expressos em amostras tumorais. A barra superior em escala verde-cinza compara a conectividade local do transcrito não codificador com a mediana da conectividade local dos transcritos codificadores nesta via, em branco os que tem a conectividade na via igual à mediana, em verde os com conectividade local superior.

UCA1 e LINC02577 ambos DE, foram localizados no *cluster* verde, o segundo mais bem correlacionado com fenótipo tumoral, esse por sua vez é enriquecido para a “via de reparo de DNA” (figura 24).

Figura 24 - *Heatmap* do módulo de coexpressão enriquecidos para a via de Reparo de DNA (*DNA repair*) com os lncRNAs DE do módulo.

REPARO DE DNA



As cores representam a distância na rede entre pares de genes, sendo sempre um lncRNA e um membro da via de Reparo de DNA (coding-RNA nas linhas e ncRNAs nas colunas, sendo pretos os mais próximos e brancos os mais distantes). Todos os não

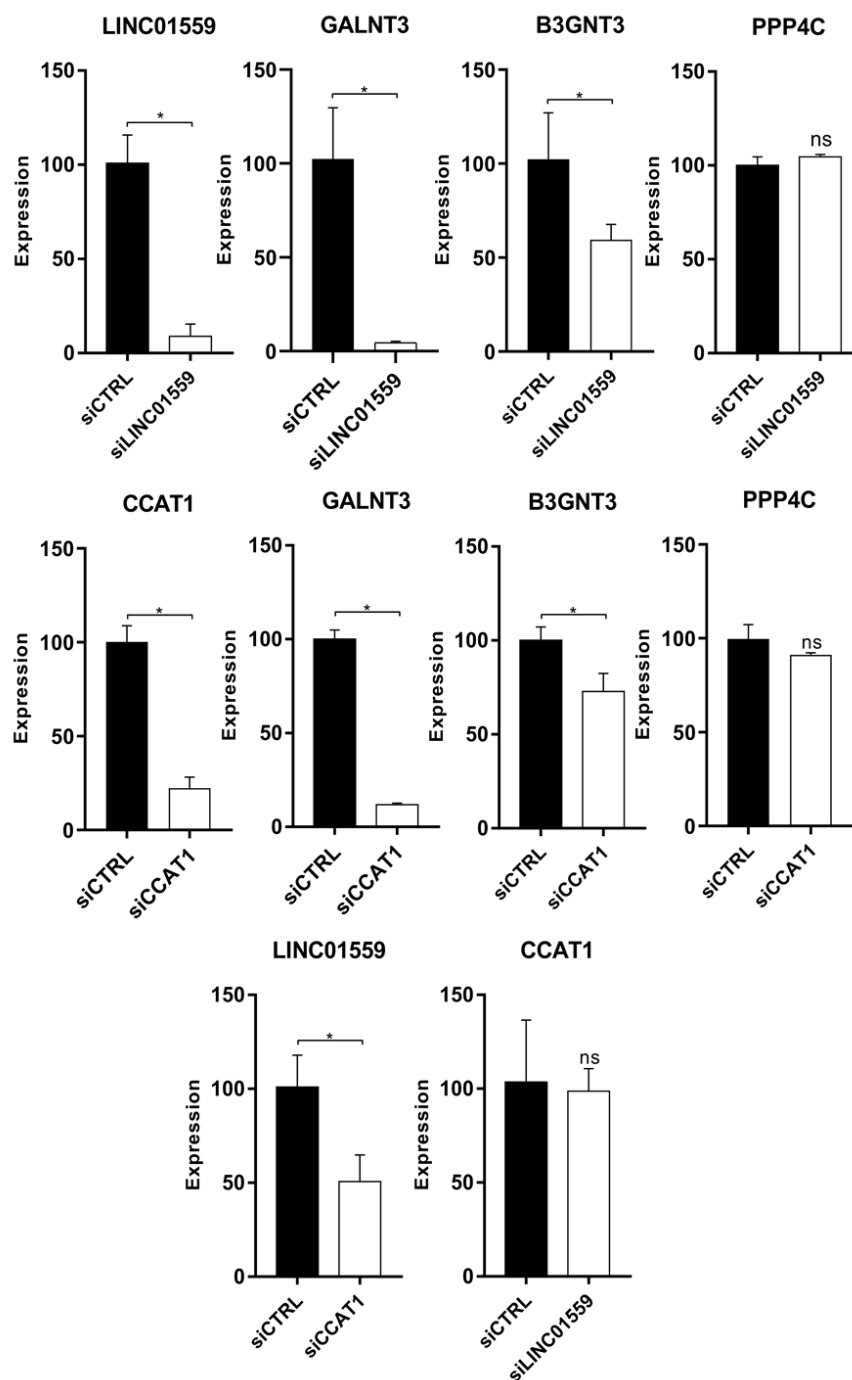
codificadores mostrados são diferencialmente expressos, barras laterais indicam em azul os que estão mais expressos em amostras controle e em vermelho os mais expressos em amostras tumorais. A barra superior em escala verde-cinza compara a conectividade local do transcrito não codificador com a mediana da conectividade local dos transcritos codificadores nesta via, em branco os que tem a conectividade na via igual à mediana, em verde os com conectividade local superior.

4.9. LINC01559 regulado pela expressão de CCAT1, regula genes da via de O-glicosilação e está enriquecido em CSCs

As correlações observadas no módulo de co-expressão de O-Glicosilação de Mucinas (Figura 23 item 4.8) para o LINC01559 foi validada em AsPC-1. No contexto de silenciamento de LINC01559 por siRNA foram investigados como se comportam as expressões relativas de *GALNT3* e *B3GNT3*, genes codificadores agrupados no mesmo *cluster* enriquecido para via de O-glicosilação de mucinas, como controle negativo desse sistema foi utilizado *PPP4C*, um gene codificador não implicado nesse módulo de co-expressão. O resultado observado foi que existe uma significativa correlação positiva entre os níveis de *LINC01559*, *B3GNT3* e *GALNT3*, enquanto com o gene controle *PPP4C*, o qual não está no mesmo módulo de correlação, não teve variações significativas dos níveis de expressão (Figura 25). CCAT1 também está localizado no mesmo módulo de de co-expressão de O-Glicosilação de Mucinas, em AsPC-1 foi silenciado eficientemente e nesse contexto *B3GNT3* teve o mesmo comportamento observado em *knockdown* de *LINC01559*, porém *LINC01559* também teve sua expressão diminuída em *knockdown* de *CCAT1*, e

o contrário não foi observado, os níveis de expressão de CCAT1 não foram significativamente alterados em *knockdown* de LINC01559.

Figura 25 – Expressão relativa de genes codificadores da rede de co-expressão do módulo de O-Glicosilação em *knockdown* de LINC01559 e CCAT1.

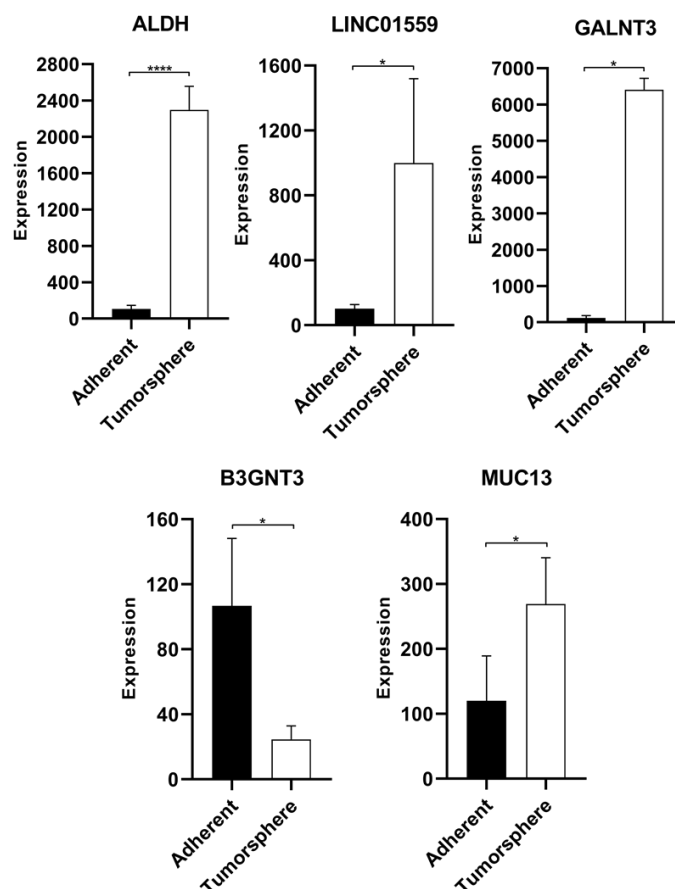


Em contexto de *knockdown* de LINC01559 (87% de silenciamento) por siRNA existe uma correlação direta com os níveis de expressão GALNT3 (92% de redução da expressão relativa) e B3GNT3 (36% de redução da expressão relativa), enquanto não

foi observada variações nos níveis do controle negativo PPP4C por RT-qPCR. Em contexto de *knockdown* de CCAT1 observa-se uma correlação positiva com a expressão de LINC01559 (42% de redução da expressão relativa), GALNT3(81% de redução da expressão relativa) B3GNT3 (31% de redução da expressão relativa), entretanto não há alterações significativas da expressão de LICN01559 em *knockdown* de CCAT1, PPP4C também foi utilizada como controle negativo. (* P-valor $\leq 0,05$).

A via enriquecida no *cluster* co-expressão “O-Glicosilação de Mucinas” contém um grande número genes para O-glicosiltransferases. Dessas GALNT3 e B3GNT3 estão implicadas na auto renovação de células-tronco tumorais (CITs) do câncer pancreático (Barkeer et al. 2018). A mensuração por RT-qPCR confirmou o enriquecimento significativo para LINC01559, GALNT3 e MUC13 (gene implicado no módulo de co-expressão) em tumoresferas quando comparados com células aderentes, B3GNT3 por outro lado demonstrou estar significativamente enriquecida na população de células aderentes de AsPC-1 (Figura 26). A expressão de ALDH foi aumentada em tumoresferas, demonstrando um enriquecimento de CITs nessa população.

Figura 26 – Expressão relativa de *LINC01559*, *GALNT3*, *B3GNT3* e *MUC13* em tumoresferas comparadas a células aderentes de AsPC-1 por RT-qPCR.



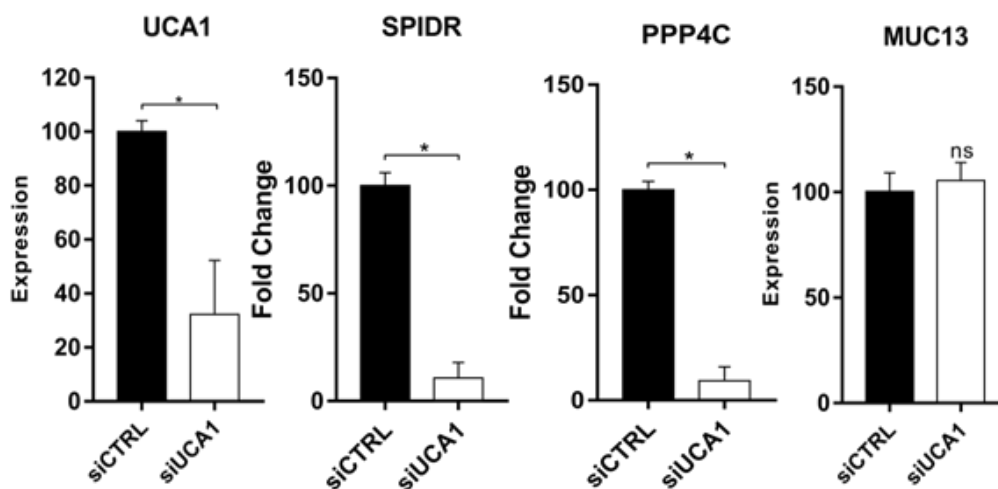
O controle ALDH é um marcador para CSCs, enriquecimento de CITs. Foi observado que ele está 2330% aumentado em tumoresferas. 1060% *LINC01559*, 6332% *GALNT3*, e 276% *MUC13* estão enriquecidos em tumoresferas enquanto *B3GNT3* a expressão relativa reduzida em tumoresferas (69% de redução da expressão relativa) (* P-valor \leq 0,05; *** P-valor \leq 0.001).

4.10. UCA1 envolvido com reparo de DNA

As análises pelo WCGNA localizaram o lncRNA DE em PDAC o UCA1 no *cluster* cujo módulo de co-expressão foi enriquecido para via de reparo de DNA (Figura 23 item 4.8). A correlação entre o lncRNA *UCA1* e os genes codificadores implicados na rede de co-expressão foi confirmada pela correlação positiva observada nas expressões dos genes *SPDR* e *PPP4C* em contexto de silenciamento de *UCA1* por siRNA. A seleção de *SPDR* e *PPP4C* em detrimento dos outros genes implicados na rede deu-se por observar-se que se trata dos

genes codificadores mais bem correlacionados da rede e cuja expressão é aumentada em PDAC. Por RT-qPCR foi confirmado que a diminuição da expressão relativa de UCA1 foi acompanhada pela significativa diminuição dos níveis de *SPIDR* e *PPP4C* (Figura 27), enquanto nenhuma alteração significativa foi observada nos níveis de expressão de *MUC13*, gene não implicado nessa rede que foi utilizado como controle.

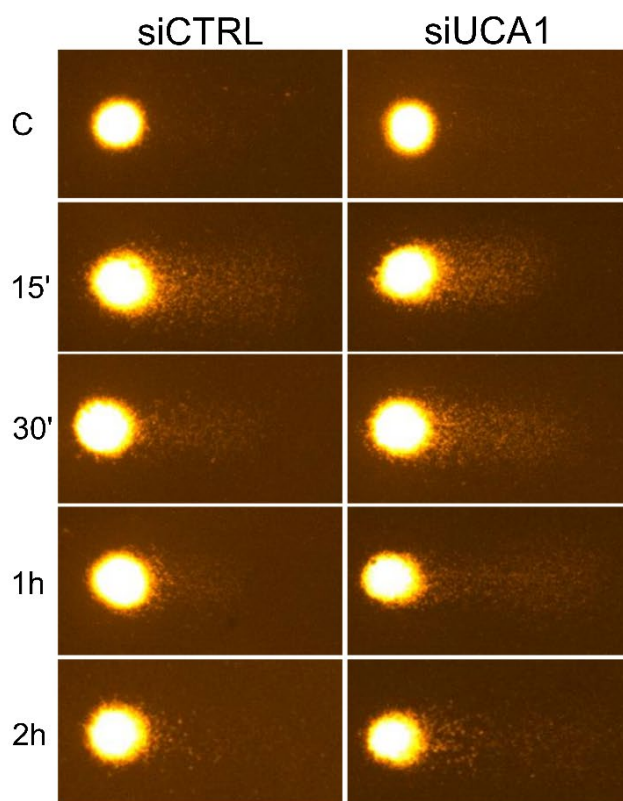
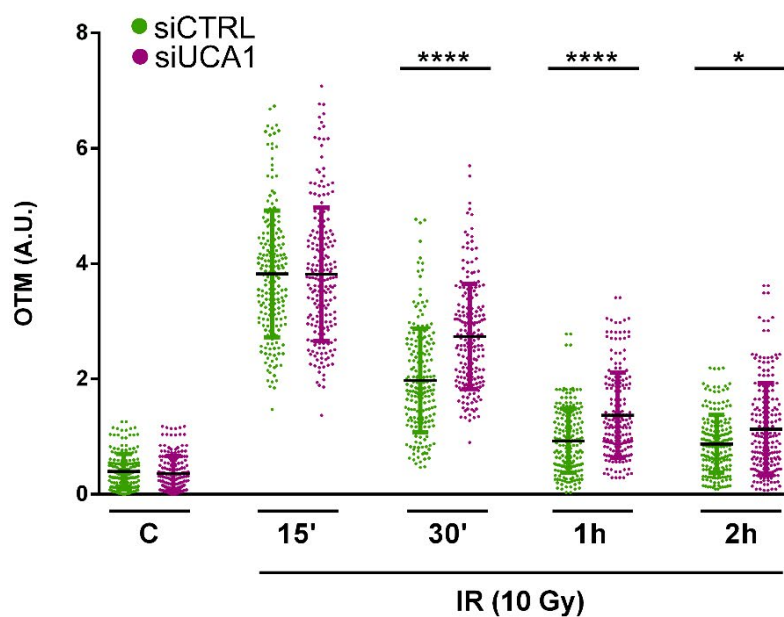
Figura 27 – Expressão relativa dos genes codificadores do módulo de Reparo de DNA *knockdown* de UCA1.



No silenciamento transiente por siRNA de UCA1 por 96h observa-se por RT-qPCR que as expressões de *SPIDR* e *PPP4C* são reduzidas significativamente em relação a condição controle. 89,33% de silenciamento para *SPIDR*; 90,66% para a *PPP4C* quando o *UCA1* é silenciado em 77,66%. Não foram observadas variações significativas de *MUC13*, um mRNA que não está implicado nesse módulo de co-expressão, como controle negativo. (* P-valor \leq 0,05).

Os genes *SPIDR* e *PPP4C*, cuja a expressão foi confiadamente regulada por *UCA1* em AsPC-1 no ensaio acima, são descritos na literatura como envolvidos no processo de reparo de quebra de fita dupla de DNA (DSB) logo foi verificado que especificamente *UCA1* altera a velocidade da capacidade de reparo de DNA em contexto de DSB provocadas por radiação γ (Figura 28).

Figura 28 – Ensaio cometa com *knockdown* de UCA1



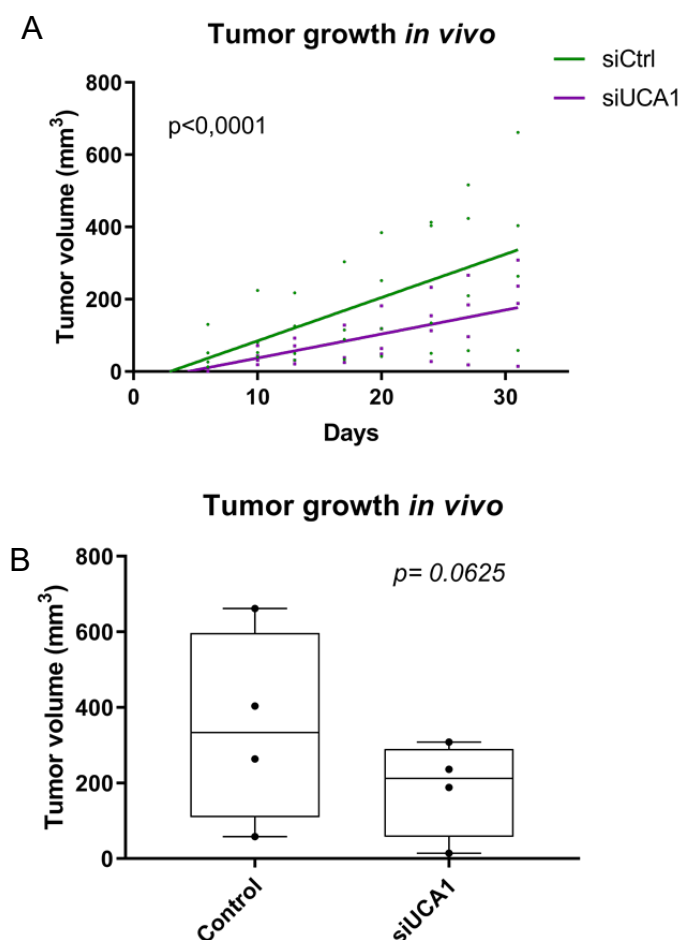
As células AsPC-1 transfectadas com siRNAs UCA1 ou siRNAs de controle foram irradiadas por radiação gama e os níveis relativos de dano ao DNA em diferentes momentos (15, 30, 60, 120 min) foram monitorados por ensaio cometa alcalino, expresso em unidades OTM (Olive Tail Moment). "C" refere-se ao valor de OTM antes da irradiação. Dados de 50 núcleos medidos em três experimentos replicados independentes são mostrados. Diferenças significativas foram detectadas usando ANOVA (*, $P \leq 0,05$) (****, p -valor $\leq 1 \times 10^{-3}$). Abaixo estão imagens representativas dos

núcleos e caudas observadas em cada uma das condições analisadas pelo experimento.

4.11. UCA1 diminui progressão tumoral em xenoenxertos

A progressão tumoral *in vivo* por xenoenxerto de derivado de linhagem AsPC-1 em camundongos foi acompanhada por 31 dias, afetada pela expressão de UCA1. Os tumores desenvolvidos a partir da injeção subcutânea de AsPC-1 transfectadas com siUCA1 tiveram uma progressão tumoral significativamente menor, chegando ao final de 31 dias com volumes tumorais 46,15% menores quando comparados os tumores knockdown de UCA1 com tumores de AsPC-1 transfectadas com siCTRL (Figura 29).

Figura 29 – Crescimento tumoral em *knockdown* de UCA1



Xenotumores derivados de linhagem AsPC-1 em camundongos. (A) A reta verde: regressão linear dos volumes dos tumores controles (n=4) 1x10⁶ células AsPC-1

(transfectadas previamente com siCTRL) administradas no subcutâneo (s.c.) dos flancos dos animais. A reta roxa: regressão linear dos volumes dos tumores *knockdown* para UCA1(siUCA1) (n=5). O dia da injeção s.c. foi considerada o tempo 0, os volumes tumorais foram monitorados a cada 3 dias para acompanhar a velocidade de crescimento dos tumores xenográficos. (análise de covariância - ANCOVA, $P \leq 0.0001$) (que compara as inclinações das retas). (B) Volumes tumorais dos controles (siCTRL) e *knockdown* para UCA1(siUCA1) ao final do período de 31 dias após a injeção s.c. O volume médio final foi de 346,71 mm³ para o tumor siCTRL e 186,7 mm³ para o tumor siUCA1 (*Mann Whitney*, P-valor = 0,0625).

5. Discussão

O câncer de pâncreas (PC) tem o pior resultado de sobrevivência quando comparada a qualquer tipo de câncer humano, com projeções de vir a ser a segunda maior causa de morte por câncer até 2030 (Lola Rahib et al. 2014; RL Siegel 2017), esse quadro demonstra a necessidade de estudos sobre o PC e principalmente sobre o tipo mais comum de PC, o PDAC. A abordagem transcriptômica tem se demonstrado uma ferramenta poderosa na estudo do PDAC (Puleo et al. 2018). Vários estudos exploraram o transcriptoma de amostras cirúrgicas de PDAC (B. J. B. Raphael et al. 2017; Moffitt et al. 2015; Collisson et al. 2019) utilizam o método de bibliotecas de seleção de transcritos poliadenilados. Bibliotecas de RNA total (depleção do rRNA) são mais enriquecidas para lncRNAs e pequenos RNAs quando comparada a outra estratégia, apesar disso fazer com que haja uma maior concentração de sequências intrônicas de mRNAs imaturos processados (S. Zhao et al. 2018), por isso nas análises foram adotados filtros estritos que eliminaram das análises transcritos raros (FPKM < 0.01, transcritos monoexônicos, e filtros de tamanho de éxons não anotados em nenhum banco de dados (Tabela 8). Com isso chegou-se ao conjunto de 62.474 transcritos reconstruídos em nossa montagem. Desses 6.710 são transcritos novos (não-anotados na referência do GENCODE) nunca descritos em PDAC. Esse resultado levanta a hipótese de a descoberta desses milhares de novos transcritos, frente a outros trabalhos, se deve a maioria dos trabalhos sobre o transcriptoma de PDAC ter investigado predominantemente bibliotecas enriquecidas de RNAs poliA+. Sendo que é inédito a abordagem desse trabalho em avaliar o transcriptoma pelo método de

RNAseq de RNA total em amostras cirúrgicas pareadas de PDAC e tecido não-tumoral adjacente. Corrobora hipótese de que este estudo explorou a fração não-poliadenilada (poliA-) do transcriptoma a análise realizada pelo prof. Eduardo Reis que demonstra que nossa reconstrução é mais enriquecida para RNAs conhecidamente poliA- (X.-O. Zhang et al. 2014; Yin, Chen, and Yang 2015), nessa análise foi constatada o enriquecimento desses genes (*RPPH1*, *TERC* e várias histonas) nos nossos dados comparados aos dados de RNAseq do TCGA (Anexo 2, figura suplementar 3). A robustez da reconstrução é corroborada pela: constatação de que uma parte considerável dos transcritos não anotados no GENCODE (figura 9) serem anotados no catálogo do banco de dados independente do miTranscriptome (Iyer et al. 2015); e também por análises comparativas que demonstram que os novos transcritos reconstruídos compartilham das mesmas características que os conhecidos. O conjunto das novas formas de *splicing* e os novos lncRNAs reconstruídas possuem a mesma distribuição que as populações anotadas no GENCODE. Quando comparados por classes: conjuntos de transcritos novos intergênicos com lncRNAs anotados no GENCODE, novas formas de *splicing* com mRNAs anotados no GENCODE, novos transcritos antisense com RNAs antisense anotados no GENCODE, verifica-se que eles tem os mesma distribuição de características estruturais (abundância, tamanho do transcrito, número de éxons por transcrito) e características funcionais (marcas regulatórias de promotor e *enhancer*) sendo que essas distribuição são todas diferentes de distribuições ao acaso (Anexo 2 - figura 2). Análises não supervisionadas (Anexo 2, figura suplementar 1) e análise supervisionadas (Anexo 2, Figura 4) de cluster hierárquico mostram que as expressões dos transcritos têm assinaturas que distinguem as amostras tumorais das não tumorais. Toas essas análises demonstram que os transcritos reconstruídos e filtrados em nossa análise são “*bona fide*”.

Diversos lncRNAs anotados e novos tem potencial translacional prognóstico, avaliando a expressão nos casos do TCGA/ICGC e a caracterização funcional realizada nesse trabalho mostra que ao menos alguns deles tem impacto em fenótipos malignos. Os mecanismos através dos quais lncRNAs exercem suas funções regulatórias ainda são desconhecidos e por isso foi utilizada a heurística da abordagem da “culpa por associação” (Lamere and Li 2019; Lefever et al. 2017), afim de inferir a esses lncRNAs associações com

funções biológicas. Neste trabalho a abordagem de “culpa por associação” foi realizada pela construção de redes de co-expressão (RCEX) utilizando-se o WCGNA (Tang et al. 2018), que identificou *clusters* nos quais os transcritos correlacionam-se pela expressão gênica (Anexo 2 – figura suplementar 7). Uma vez correlacionados, a inferência de função é feita pelo enriquecimento de vias para cada módulo, observado nos 22 módulos de co-expressão resultantes da análise (Figura 22). Essa análise teve resultados coerentes, pois o *cluster* roxo, mais bem correlacionado com o fenótipo tumoral, está enriquecido para vias classicamente descritas como relevantes em tumores, esse *cluster* em específico é enriquecido para “glicosilação de mucinas” PDAC (Kaur et al 2013). Corrobora a robustez dessa análise o fato do cluster enriquecido para “secreção pancreática” ser o menos correlacionado com fenótipo tumoral, fato que pode ser explicado pela perda de função endócrina e exócrina pancreática no tecido tumoral. A partir da análise por RCEX foram verificados módulos enriquecidos para vias de reparo de DNA e O-glicosilação de mucinas (Figuras 23 e 24).

Todos os 6 (seis) novos transcritos selecionados a partir dos dados de RNAseq tiveram a existência validada (Figura 10) por PCR *end point* e a DE validada por RT-qPCR de amostras cirúrgicas (Figura 12), T e N, e amostra de PDX, confirmando-se a expressão aberrante em contexto tumoral (T e PDX) quando comparado ao tecido não tumoral adjacente (N). Os novos intergênicos TCONS00084995, TCONS00009076, TCONS00087289, TCONS00006750, TCONS00059572, tem sua estrutura corroborada pelo miT, já TCONS000036574 aparece unicamente em nossa reconstrução, entretanto até o presente momento nenhum desses transcritos foram validados por outros grupos. As expressões relativas desses foram medidas nas linhagens de PDAC disponíveis no laboratório (Figura 11) e em linhagem derivada de PDX, observou-se que os dois novos intergênicos, TCONS00059572 e TCONS000036574, foram detectados em níveis significativos apenas na linhagem de PDX. Esses resultados demonstram o viés característico do cultivo de linhagens celulares em 2D, em que cada linhagem recuperara apenas parte distinta do comportamento transcricional quando comparadas a amostra original tridimensional. Possivelmente a linhagem derivada de PDX recupera ainda, por ter pouquíssimas passagens, a organização da superfície celular que ocorre na cultura 3D e no tumor *in situ* com isso algumas vias de sinalização alternativas

que podem não estar disponíveis na cultura 2D (Pickl and Ries 2009). E essas vias podem ser as vias relacionadas com a expressão de TCONS00059572 e TCONS00036574. Desse conjunto três deles, TCONS00036574, TCONS00085964 e TCONS00006750, foram correlacionados pela análise de redes de co-expressão dentro do módulo *O-glicosililation* (Figura 23), indicando uma possível associação regulatória com os genes codificadores agrupados na rede constituída por aquele módulo, esse módulo tem sobreposição pelas O-glicosiltransferases GALNT3 e B3GNT3 com a manutenção de CSCs em PDAC (Barkeer et al. 2018). A análise de sobrevida mostrou que para dois deles TCONS00085964 e TCONS00036574 existe um impacto na sobrevida global dos pacientes (Figura 13). Sendo que uma maior expressão de TCONS00085964 está correlacionada com uma diminuição da sobrevida, e uma *hazard ratio* (HR) de 1,6 e um *p*-valor marginal de 0,085. Ao passo que uma maior expressão de TCONS00036574 tem uma (HR) de 0,6 o *p*-valor 0,044 está associada com o aumento da sobrevida global do paciente, demonstrando o potencial translacional desses novos intergênicos revelados nesse trabalho. A análise de sobrevida desses grupos de novos intergênicos foi feita a partir dos dados brutos do banco de dados de RNAseq do *International Cancer Genome Consortium* (ICGC) com a nossa montagem do transcriptoma e observamos maior significância estatística quando utilizado esse banco de dados (PACA-AU), esses dados foram obtidos de RNAseq de bibliotecas de RNA total, como as deste trabalho.

As novas AS validadas (Figura 14) MMP14 (TCONS00027381), CAPN8 (TCONS00006585), LIF (TCONS00057528) e OCT3, ou SLC22A3, (TCONS00074972), todas constam no catálogo do miT, são formas de *splicing* não catalogadas pelo GENCODE, até o presente não foram encontrados registros de validações em nível de proteína dessas AS. Os quatro genes canônicos são sabidamente envolvidos com câncer (Jiang et al. 2015; D.-T. Chen et al. 2015; Hata et al. 2016; Y. Shi et al. 2019; Fu et al. 2017; Hsu et al. 2017). MMP14 é uma metaloproteinase (MMP) responsável pelo remodelamento da matriz extracelular. As MMPs foram propostas como possíveis alvos terapêuticos, e foi relatado que podem influenciar o microambiente tumoral levando a um maior crescimento, angiogênese e metástase (Cathcart, Pulkoski-Gross, and Cao 2015). CAPN (Calpaínas) é uma

família de proteases dependentes de cálcio que regula vários processos celulares clivando substratos (Hata et al. 2016), e já foi correlacionada a expressão de CAPN8 com pior prognóstico em PDAC (D.-T. Chen et al. 2015). Recentemente Bressy e colaboradores encontraram uma sobre-expressão de LIF em tecidos de câncer pancreático, e detectaram sua presença em soro de pacientes e de camundongos com a doença mas não nos indivíduos saudáveis, indicando que poderia servir como biomarcador (Bressy et al. 2018). As 4 (quatro) AS validadas tem expressão aumentada em PDAC e são formas predominantes de expressão dos genes em PDAC, sendo potenciais antígenos específicos de tumor (TSAs) (Smith et al. 2019).

Este trabalho está no seguimento de outros do grupo de pesquisa como o Tahira et al., a abordagem para investigar lncRNAs implicados em câncer de pâncreas, PDAC, a partir de lncRNA com expressão aberrante no tumor (Tahira et al. 2011b; Reis and Verjovski-Almeida 2012b). Constatou-se que todos os lncRNAs selecionados e que foram validados (Figura 15) estão envolvidos no fenótipo tumoral de PDAC, LINC01559; LINC01133; CCAT1; LINC00920 e UCA1 (Figuras 19, 20 e 21).

O LINC01614 foi descrito como diferencialmente expresso em carcinoma epidermóide escamoso de esôfago (W. Wang et al. 2018), em adenocarcinoma de pulmão, câncer de pulmão de não pequenas células e câncer de mama foi também detectado em linhagens desses cânceres (A. Liu et al. 2018; Vishnubalaji et al. 2019) . Neste trabalho a expressão de LINC01614 foi detectada como aberrante em PDAC pelo RNAseq, na validação por RT-qPCR apenas detectada em amostras de PDAC cirúrgica, não sendo detectados significativamente em amostras não tumorais cirúrgicas e no conjunto de amostras PDX (Figura 15). Esse resultado sugere que em PDAC o LINC01614 é expresso apenas em contexto de TME humano, podendo ser expresso por células específicas do TME ou ser expresso pelas células tumorais por vias de sinalização que os modelo experimental adotado, PDX, não foi capaz de recapitular mesmo com a presença do estroma do camundongo hospedeiro. Por isso, nesse trabalho, o LINC01614 não foi avaliado quanto a seu envolvimento com migração, invasão e proliferação por não ser expresso por nenhum dos modelos experimentais avaliados, cultura de linhagens celulares de PDAC em 2D, bem como em PDX.

LINC01133, envolvido com migração, invasão e progressão tumoral neste trabalho (Figuras 19, 20 e 21) já foi descrito por ter altos níveis de expressão em tipos de câncer como pulmonar (NSCLC), tumores de cólon (CRC) e baixos níveis em tumor gástrico (GC). Nota-se na literatura que um comportamento de funções contexto-dependentes, sendo descrito como supressor tumoral em CRC e GC, e tumorigênico em NSCLC. O LINC01133 interage com fator de *splicing* SRSF6, o qual tem um papel de regulador de, agindo como ligantes alternativo em SRSF6. Relatado como tumorigênico em NSCLC, o mecanismo consiste em basicamente a interação do LINC01133 com EZH2 e LSD1, com consequente inibição da expressão de P21, KLF2 e E-Caderina. Nesse caso foi observado que o silenciamento de LINC01133 em linhagens A549 levou a uma diminuição do crescimento (*in vivo*), migração e invasão. No contexto de PDAC foi demonstrado que LINC01133 atua de forma tumorigênica através do envolvimento com a metilação do promotor de DKK1 e a ativação da via de sinalização da Wnt (Weng et al. 2019). Um outro dado na literatura mostra que o LINC01133 atua como supressor tumoral em Câncer Gástrico (GC), diminuindo a progressão e metástase agindo como ceRNA para o miR-106a-3p o qual por sua vez regula a expressão de APC consequentemente a via Wnt/ β -catenina (Zang et al. 2016; Kong et al. 2016; Wan et al. 2019; X.-Z. Yang et al. 2018).

O Transcrito Associado ao Câncer de Cólon 1 (CCAT1), envolvido com proliferação, migração e invasão nesse trabalho (Figuras 19, 20 e 21), tem seu *locus* na vizinhança de um fator de transcrição bem conhecido, c-Myc, e estudos mostram que a expressão de CCAT1 é ativada por c-Myc, o aumento da expressão de CCAT1 contribui para a tumorigênese, processo metastático em do CCR, GC e PDAC. Porém o mecanismo pelo qual CCAT1 atua não foi totalmente elucidado, na literatura encontra-se o seu envolvimento por ter um papel de esponja de let-7 em câncer hepatocelular (HCC), sendo por isso envolvido com o aumento da progressão tumoral e marcador de prognóstico nesse câncer. Nos nossos dados CCAT1 foi localizado nas redes de co-expressão dentro do módulo enriquecido para via de O-glicosilação de mucinas (Figura 23), e demonstrou ser um regulador da expressão de LINC01559, GALNT3 e B3GNT3 (Figura 25), pois em *knockdown* de CCAT1 houve uma correlação positiva com a expressão de LINC01559 B3GNT3, entretanto não

observa-se alterações significativas da expressão de LINC01559 em *knockdown* de CCAT1, nem no controle negativo PPP4C, isso aponta para hipótese de um eixo que tem CCAT1 como um elemento regulador a *upstream* de LINC01559.

LINC00920, neste trabalho foi observado estar envolvido com migração e invasão (Figuras 19 e 20), em pelo modelo experimental de AsPC-1 em 2D não foi observada resultados significados quanto proliferação, pesquisas com outros modelos experimentais são necessárias para explorar mais a implicação de LINC00920 em PDAC. Foi descrito com expressão aberrante, proliferação tumoral e associado com diminuição da sobrevida em CRC, pelo mecanismo no qual LINC00920 interage com sítios de ubiquitinação do IGF2BP2 e previne sua degradação (Yun Wang et al. 2019).

O LINC02577 (Figura 15) assim denominado na versão GRCh38 do genoma humano utilizada como referência do GENCODE atualmente, esse gene era denominado RP5-884M6.1 na versão anterior GRCh37. Foi descrito como sendo superexpresso em câncer de nasofaringe, superexpresso em carcinoma oral de células escamosas (Y. Song, Pan, and Liu 2019), também foi descrito como sendo uma possível assinatura para adenocarcinoma de estômago (J. Yang and Song 2019). Outros o trabalhos mostram que a expressão de LINC02577 é afetada por irradiação por UVB além de interagir com a cromatina em linhagem K562 (Wen et al. 2016; Cai et al. 2016). Esse dado corrobora os resultados das análises de rede de co-expressão desse estudo no qual ele foi localizado no módulo enriquecido para vias de reparo de DNA, uma vez que a principal via estudada por irradiação com UVB é a via de reparo de DNA.

LINC01559 demonstrou ser envolvido com proliferação, migração e invasão nesse trabalho (Figuras 19, 20 e 21) além disso tem potencial translacional detectado pela análise de sobrevida (Figura 17). LINC01559 foi recentemente descrito como sendo implicado na diminuição da sobrevida em adenocarcinoma de pulmão (X. Shi et al. 2018), bem como foi descrito como uma potencial assinatura sérica de câncer renal (B. Chen et al. 2018). Porém nenhum desses trabalhos investigou qualquer função biológica para o LINC01559. Nas análises de redes de co-expressão e enriquecimento de vias (Figura 22) os candidatos mais DE apareceram no módulo de *O-linked* glicosilação (Figura 23): LINC01559, CCAT1 e os novos intergênicos TCONS00036574, TCONS00085964 e TCONS00006750. Feito o silenciamento de LINC01559 por

siRNA em linhagem AsPc-1, os níveis de expressão de CCAT1, dos genes codificadores B3GNT3 e GALNT3 tiveram uma correlação direta na expressão gênica, ou seja tiveram uma diminuição em relação a situação controle, e o gene MUC13 uma correlação inversa, ou seja, quando silenciado LINC01559 houve um aumento da expressão de MUC13. Porém as correlações encontradas experimentalmente em *knockdown* diferem-se em alguns casos das observadas nas análises do WCGNA, é necessário fazer aqui duas considerações, o RNAseq é uma análise do transcriptoma, que contém populações de células diferentes como PDAC e estroma, logo o que fora observado nos módulos de co-expressão pode ter sido um comportamento de células do TME ou mesmo um comportamento das células tumorais em *crosslink* com TME, ou ainda, que a linhagem 2D utilizadas como modelo experimental não tenha recapitulado os níveis de sinalização *in situ*. A alteração dos níveis de enzimas GALNT é associado com muitos tipos de câncer, e lncRNAs superexpressos em câncer tem mecanismos de esponjas de miRNAs que alvejam transcritos GALNTs (Shan et al. 2018). As N-acetilgalactosaminiltransferases (GALNTs) são uma família do polipeptídeo N-acetilgalactosamina (GalNAc)-transferases, as quais tem por função iniciar a O-glicosilação no aparelho de Golgi, nesse processo quimicamente ocorre a catalisação da reação de adição de resíduos glicosídicos de GalNAc em sítios de resíduos de serina e treonina na proteína alvo (Bennett et al. 2012; Gerken et al. 2006). LINC01559 tem papel regulatório de B3GNT3 e GALNT3, possivelmente por esponja de miRNAs alvos de ambas ou por recrutar para esses genes fatores de transcrição. Análises *in silico* serem inconclusivas quando a possíveis sítios de miRNAs compartilhados por LINC01559 e B3GNT3 e GALNT3, entretanto fazem-se necessárias comprovações experimentais. B3GNT3 e GALNT3 são glicosiltransferases implicadas em câncer, e foi demonstrado seu papel na glicosilação e mediação do crescimento e da manutenção de CSCs em PDAC (Barkeer et al. 2018).

O lncRNA UCA1 demonstrou ser envolvido com proliferação, migração e invasão nesse trabalho (Figuras 19, 20 e 21) além disso tem potencial translacional detectado pela análise de sobrevivência (Figura 17). UCA1 foi observado com expressão aumentada em diversos tipos de tumores, incluindo câncer pancreático favorecendo a proliferação celular, migração e invasão (P. Chen et al. 2016; X. Zhang et al. 2017; Y. Zhou et al. 2018; M. Zhang et al. 2018).

Zhang et. al. sugere que *UCA1* participa na via de sinalização Hippo interagindo com *MOB1*, *Lats1* e *YAP* promovendo a migração celular (M. Zhang et al. 2018). Zhou et. al. fizeram análises de microarray em 36 tecidos tumorais e 16 normais, determinando que *UCA1* atua regulando negativamente *miR-96* e positivamente *FOXO3* (Y. Zhou et al. 2018); enquanto que Zhang et. al. sugerem que *UCA1* é um oncogene com função de esponja do microRNA *miR-135a*, regulando o crescimento e metástase do câncer pancreático (X. Zhang et al. 2017). A sobre-expressão de *UCA1* também foi correlacionada com resistência a quimioterápicos como gemcitabina, cisplatina, tamoxifen, 5-FU, imatinib e EGFR-TKIs, enquanto que a inativação de *UCA1* restabelece a sensibilidade às drogas (H. Wang et al. 2017). Todos esses trabalhos destacam a importância de *UCA1* como biomarcador diagnóstico e prognóstico, assim como também como possível alvo terapêutico em câncer pancreático (H. Wang et al. 2017). *UCA1* também foi descrito como sendo tendo envolvimento com metabolismo de glutamina em câncer de bexiga pela via mTOR (H.-J. Li et al. 2015) e metabolismo de glicose em câncer de bexiga pelo eixo miR-16-GLS2 (Zhengkun Li et al. 2014). Em câncer de bexiga *UCA1* é envolvido com migração e invasão pelo mecanismo de *UCA1/let-7e/HMGA2* (Yi Cheng et al. 2019). Neste trabalho *UCA1* foi localizado pelas redes de co-expressão em um módulo enriquecido para via de reparo de DNA, no qual ele é altamente correlacionado com transcritos de mRNA, no topo da lista de correlação estão SPIDR e PPP4C. A proteína SPIDR (proteína *scaffolding* envolvida no reparo de DNA) atua como o elo entre a helicase BLM RecQ Like e o maquinário de reparo por recombinação homóloga (HR). É proposto que SPIDR interage independentemente com BLM e RAD51 e promove a formação de um complexo de importância biológica contendo BLM / RAD51 (W. Li et al. 2013). A proteína PPP4C por sua vez é descrita como uma *druggable* fosfatase sensível à Fostriecina e demonstram que a supressão da PPP4C desencadeia apoptose (Theobald et al. 2013). Foi confirmado que o *knockdown* de *UCA1* leva a diminuição significativa da expressão de SPIDR e PPP4C, somando-se ao fato que foi confirmado que o *knockdown* de *UCA1* prejudica significativamente o reparo de DNA após quebras induzidas por radiação- γ em células de AsPC-1 (Figura 28), demonstrando que a cinética de reparo de células *knockdown* de *UCA1* tem significativamente menor velocidade de reparo menor do que o grupo transfectado com siCTRL,

sendo que ambos os grupos partem confiadamente dos mesmos níveis de fragmentação do DNA genômico, além disso o *knockdown* de UCA1 apresentou uma redução da velocidade de crescimento tumoral (Figura 28) como já fora observado na literatura (Hu et al. 2016) corroborando nossos dados.

6. Conclusão

O transcriptoma de bibliotecas de RNA total pareadas por RNAseq revelou com maior resolução características moléculas de PDAC não observadas até o momento pela literatura. Foi gerado um transcriptoma composto por 41.341 transcritos conhecidos pela referência do GENCODE e 6.710 transcritos novos quando comparados a referência do GENCODE. Desse conjunto foram classificados 339 como sendo novos transcritos intergênicos, 184 antisensos, e 6.187 potenciais novas variantes de *splicing*. A maioria deles possuem um baixo potencial codificador (Sosa 2019).

Dos 86 RNAs intergênicos, 6 novos intergênicos foram identificados com expressão aumentada em PDAC. Eles receberam pela montagem as seguintes identificações: TCONS00085964, TCONS00009076, TCONS00087289, TCONS00006750, TCONS00059572, TCONS000036574. Desses TCONS00085964 e TCONS00036574 tem correlação com alteração na sobrevivência de paciente, demonstrando ser um potencial translacional, como por exemplo marcador de prognóstico. Ensaios para a validação de regiões maiores desses novos transcritos estão sendo realizadas em nosso laboratório, eles tem como objetivo investigar a relevância desses lncRNAs novos no fenótipo tumoral bem como investigar o envolvimento em funções biológicas. Isso tudo demonstra que para esses novos transcritos os resultados desse trabalho abrem caminhos para estudos posteriores.

As análises de expressão diferencial permitiram encontrar 171 novas formas de *splicing*, das quais foram validadas novas AS de genes sabidamente implicados em câncer como MMP14, CAPN8, LIF e OCT3. A análise do transcriptoma não identificou alterações significativas na abundância das formas canônicas já anotadas destes genes, sugerindo que as formas de *splicing* validadas são as isoformas predominantemente expressas no PDAC.

Os lncRNAs com expressão aberrante identificados LINC01559, LINC01133, LINC02577, CCAT1, LINC00920 e UCA1 e selecionados demonstraram ter relevância no contexto de PDAC, LINC01559, LINC02577 e UCA1 tiveram as expressões em PDAC relacionadas com alterações na sobrevivência, indicando um potencial como marcador prognóstico. Os lncRNAs LINC01559, LINC01133, CCAT1, LINC00920 e UCA1 foram implicados com o fenótipo tumoral de PDAC, visto que o *knockdown* desses teve como resultado diminuição da migração e invasão. Esses resultados são corroborados pelo *knockdown* de LINC01559, LINC01133, CCAT1 e UCA1 causar uma diminuição na progressão tumoral. Relações com funções biológicas do fenótipo tumoral foram observadas para LINC01559 sendo envolvimento na regulação da expressão de mRNA de O-glicosilação, mais especificamente enzimas envolvidas com CSCs, a expressão de LINC01559 foi regulada pela expressão de CCAT1, por fim o envolvimento UCA1 com reparo de DNA e alteração do crescimento tumoral *in vivo*, a alteração da progressão tumoral *in vivo* já fora demonstrada anteriormente (Hu et al. 2016), o que corrobora o resultado desse experimento.

Dado o exposto nesse trabalho fica evidente a produção de um catálogo que interrogou mais elementos do transcriptoma de PDAC e identificou transcritos relevantes no contexto de PDAC com potenciais prognósticos e terapêuticos.

7. Referências

- A Scarpa, FX Real, C Luchini. 2018. "Genetic Unrelatedness of Co-Occurring Pancreatic Adenocarcinomas and IPMNs Challenges Current Views of Clinical Management." *Gut* 67: 1561–63.
- Aaij, C., and P. Borst. 1972. "The Gel Electrophoresis of DNA." *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 269 (2): 192–200. [https://doi.org/10.1016/0005-2787\(72\)90426-1](https://doi.org/10.1016/0005-2787(72)90426-1).
- Alizadeh, Ash A., Michael B. Eisen, R. Eric Davis, Chi Ma, Izidore S. Lossos, Andreas Rosenwald, Jennifer C. Boldrick, et al. 2000. "Distinct Types of

- Diffuse Large B-Cell Lymphoma Identified by Gene Expression Profiling.” *Nature* 403 (6769): 503–11. <https://doi.org/10.1038/35000501>.
- Amin, Noorul, Annette McGrath, and Yi-Ping Phoebe Chen. 2019. “Evaluation of Deep Learning in Non-Coding RNA Classification.” *Nature Machine Intelligence* 1 (5): 246–56. <https://doi.org/10.1038/s42256-019-0051-2>.
- Anders, S., P. T. Pyl, and W. Huber. 2015. “HTSeq--a Python Framework to Work with High-Throughput Sequencing Data.” *Bioinformatics* 31 (2): 166–69. <https://doi.org/10.1093/bioinformatics/btu638>.
- Aoki, K., Y. Ogata, and D. Shibata. 2007. “Approaches for Extracting Practical Information from Gene Co-Expression Networks in Plant Biology.” *Plant and Cell Physiology* 48 (3): 381–90. <https://doi.org/10.1093/pcp/pcm013>.
- Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. “Gene Ontology: Tool for the Unification of Biology.” *Nature Genetics* 25 (1): 25–29. <https://doi.org/10.1038/75556>.
- Bailey, Peter, David K. Chang, Katia Nones, Amber L. Johns, Ann-Marie Patch, Marie-Claude Gingras, David K. Miller, et al. 2016. “Genomic Analyses Identify Molecular Subtypes of Pancreatic Cancer.” *Nature* 531 (7592): 47–52. <https://doi.org/10.1038/nature16965>.
- Balachandran, VP. 2017. “Identification of Unique Neoantigen Qualities in Long-Term Survivors of Pancreatic Cancer.” *Nature* 551: 512–16.
- Baralle, Francisco E., and Jimena Giudice. 2017. “Alternative Splicing as a Regulator of Development and Tissue Identity.” *Nature Reviews Molecular Cell Biology* 18 (7): 437–51. <https://doi.org/10.1038/nrm.2017.27>.
- Barkeer, Srikanth, Seema Chugh, Saswati Karmakar, Garima Kaushik, Sanchita Rauth, Satyanarayana Rachagani, Surinder K. Batra, and Moorthy P. Ponnusamy. 2018. “Novel Role of O-Glycosyltransferases GALNT3 and B3GNT3 in the Self-Renewal of Pancreatic Cancer Stem Cells.” *BMC Cancer* 18 (1): 1157. <https://doi.org/10.1186/s12885-018-5074-2>.
- Basturk, Olca, Seung-Mo Hong, Laura D. Wood, N. Volkan Adsay, Jorge Albores-Saavedra, Andrew V. Biankin, Lodewijk A.A. Brosens, et al. 2015. “A Revised Classification System and Recommendations From the Baltimore Consensus Meeting for Neoplastic Precursor Lesions in the Pancreas.” *The American Journal of Surgical Pathology* 39 (12): 1730–41.

<https://doi.org/10.1097/PAS.0000000000000533>.

- Bennett, E. P., U. Mandel, H. Clausen, T. A. Gerken, T. A. Fritz, and L. A. Tabak. 2012. "Control of Mucin-Type O-Glycosylation: A Classification of the Polypeptide GalNAc-Transferase Gene Family." *Glycobiology* 22 (6): 736–56. <https://doi.org/10.1093/glycob/cwr182>.
- Berger, Adam C, Miguel Garcia, John P Hoffman, William F Regine, Ross A Abrams, Howard Safran, Andre Konski, et al. 2008. "Postresection CA 19-9 Predicts Overall Survival in Patients with Pancreatic Cancer Treated with Adjuvant Chemoradiation: A Prospective Validation by RTOG 9704." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 26 (36): 5918–22. <https://doi.org/10.1200/JCO.2008.18.6288>.
- Biankin, Andrew V., Nicola Waddell, Karin S. Kassahn, Marie-Claude Gingras, Lakshmi B. Muthuswamy, Amber L. Johns, David K. Miller, et al. 2012. "Pancreatic Cancer Genomes Reveal Aberrations in Axon Guidance Pathway Genes." *Nature* 491 (7424): 399–405. <https://doi.org/10.1038/nature11547>.
- Biswas, Prodip Kumar, Md Hafiz Sardar, Gopal Chandra Saha, Mohammed Shahadat Hossain, Mohammad Motlabur Rahman, Mostofa Kamal Chowdhury, and Tushit Paul. 2019. "Clinical Value of Serum Cancer Antigen 19-9 as a Tumor Screening Marker among Healthy Individuals." *Journal of Medicine* 20 (2): 92–94. <https://doi.org/10.3329/JOM.V20I2.42009>.
- Brand, Randall E, Markus M Lerch, Wendy S Rubinstein, John P Neoptolemos, David C Whitcomb, Ralph H Hruban, Teresa A Brentnall, Henry T Lynch, Marcia I Canto, and Participants of the Fourth International Symposium of Inherited Diseases of the Pancreas. 2007. "Advances in Counselling and Surveillance of Patients at Risk for Pancreatic Cancer." *Gut* 56 (10): 1460–69. <https://doi.org/10.1136/gut.2006.108456>.
- Bray, Freddie, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. 2018. "Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians* 68 (6): 394–424. <https://doi.org/10.3322/caac.21492>.
- Bray, Nicolas L, Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. "Near-Optimal Probabilistic RNA-Seq Quantification." *Nature Biotechnology* 34 (5):

525–27. <https://doi.org/10.1038/nbt.3519>.

- Breasted, James Henry. 1931. "The Edwin Smith Surgical Papyrus: Published in Facsimile and Hieroglyphic Transliteration with Translation and Commentary in Two Volumes." *JAMA: The Journal of the American Medical Association* 96 (18): 1534. <https://doi.org/10.1001/jama.1931.02720440082042>.
- Brett, David, Jens Hanke, Gerrit Lehmann, Sabine Haase, Sebastian Delbrück, Steffen Krueger, Jens Reich, and Peer Bork. 2000. "EST Comparison Indicates 38% of Human MRNAs Contain Possible Alternative Splice Forms." *FEBS Letters* 474 (1): 83–86. [https://doi.org/10.1016/S0014-5793\(00\)01581-7](https://doi.org/10.1016/S0014-5793(00)01581-7).
- Cai, Liuyang, Huidan Chang, Yaping Fang, and Guoliang Li. 2016. "A Comprehensive Characterization of the Function of LincRNAs in Transcriptional Regulation Through Long-Range Chromatin Interactions." *Scientific Reports* 6 (1): 36572. <https://doi.org/10.1038/srep36572>.
- Calabretta, S, P Bielli, I Passacantilli, E Piloizzi, V Fendrich, G Capurso, G Delle Fave, and C Sette. 2016. "Modulation of PKM Alternative Splicing by PTBP1 Promotes Gemcitabine Resistance in Pancreatic Cancer Cells." *Oncogene* 35 (16): 2031–39. <https://doi.org/10.1038/onc.2015.270>.
- Chakraborty, Subhankar. 2014. "In Silico Analysis Identifies Genes Common between Five Primary Gastrointestinal Cancer Sites with Potential Clinical Applications." *Ann Gastroenterol* 27 (March): 1–11. <https://doi.org/10.3233/CBM-182296>.
- Chen, Binghai, Chengyue Wang, Jin Zhang, Yang Zhou, Wei Hu, and Tao Guo. 2018. "New Insights into Long Noncoding RNAs and Pseudogenes in Prognosis of Renal Cell Carcinoma." *Cancer Cell International* 18 (1): 157. <https://doi.org/10.1186/s12935-018-0652-6>.
- Chen, Dung-Tsa, Ashley H. Davis-Yadley, Po-Yu Huang, Kazim Husain, Barbara A. Centeno, Jennifer Permuth-Wey, Jose M. Pimiento, and Mokenge Malafa. 2015. "Prognostic Fifteen-Gene Signature for Early Stage Pancreatic Ductal Adenocarcinoma." Edited by Francisco X. Real. *PLOS ONE* 10 (8): e0133562. <https://doi.org/10.1371/journal.pone.0133562>.
- Cheng, Yating, Indira Jutooru, Gayathri Chadalapaka, J Christopher Corton, and Stephen Safe. 2015. "The Long Non-Coding RNA HOTTIP Enhances Pancreatic Cancer Cell Proliferation, Survival and Migration." *Oncotarget* 6

- (13): 10840–52. <https://doi.org/10.18632/oncotarget.3450>.
- Cheng, Yi, Chunliu Huang, Yongxuan Mo, Weiwu Wu, and Lu Liang. 2019. “Long Non-Coding RNA UCA1 Regulates Tumor Growth by Impairing Let-7e-Dependent HMGA2 Repression in Bladder Cancer.” *Cancer Biomarkers*, July, 1–11. <https://doi.org/10.3233/CBM-182296>.
- Cogill, Steven B., and Liangjiang Wang. 2014. “Co-Expression Network Analysis of Human LncRNAs and Cancer Genes.” *Cancer Informatics* 13s5 (January): CIN.S14070. <https://doi.org/10.4137/CIN.S14070>.
- Collins, Francis S, Michael Morgan, and Aristides Patrinos. 2003. “The Human Genome Project: Lessons from Large-Scale Biology.” *Science (New York, N.Y.)* 300 (5617): 286–90. <https://doi.org/10.1126/science.1084564>.
- Collins, Tony J. 2007. “ImageJ for Microscopy.” *BioTechniques* 43 (1S): S25–30. <https://doi.org/10.2144/000112517>.
- Collisson, Eric A., Peter Bailey, David K. Chang, and Andrew V. Biankin. 2019. “Molecular Subtypes of Pancreatic Cancer.” *Nature Reviews Gastroenterology & Hepatology* 16 (4): 207–20. <https://doi.org/10.1038/s41575-019-0109-y>.
- Collisson, Eric A, Anguraj Sadanandam, Peter Olson, William J Gibb, Morgan Truitt, Shenda Gu, Janine Cooc, et al. 2011. “Subtypes of Pancreatic Ductal Adenocarcinoma and Their Differing Responses to Therapy.” *Nature Medicine* 17 (4): 500–503. <https://doi.org/10.1038/nm.2344>.
- Conesa, Ana, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szczęśniak, et al. 2016. “A Survey of Best Practices for RNA-Seq Data Analysis.” *Genome Biology* 17 (1): 13. <https://doi.org/10.1186/s13059-016-0881-8>.
- Connor, Ashton A., and Steven Gallinger. 2017. “Next Generation Sequencing of Pancreatic Ductal Adenocarcinoma: Right or Wrong?” *Expert Review of Gastroenterology & Hepatology* 11 (7): 683–94. <https://doi.org/10.1080/17474124.2017.1324296>.
- Corbo, Vincenzo, Giampaolo Tortora, and Aldo Scarpa. 2012. “Molecular Pathology of Pancreatic Cancer: From Bench-to-Bedside Translation.” *Current Drug Targets* 13 (6): 744–52. <http://www.ncbi.nlm.nih.gov/pubmed/22458520>.
- Croft, Larry, Soeren Schandorff, Francis Clark, Kevin Burrage, Peter Arctander,

- and John S. Mattick. 2000. "ISIS, the Intron Information System, Reveals the High Frequency of Alternative Splicing in the Human Genome." *Nature Genetics* 24 (4): 340–41. <https://doi.org/10.1038/74153>.
- Cukierman, E., R Pankov, D R Stevens, and K M Yamada. 2001. "Taking Cell-Matrix Adhesions to the Third Dimension." *Science* 294 (5547): 1708–12. <https://doi.org/10.1126/science.1064829>.
- Dalen, Floris J van, Marleen H M E van Stevendaal, Felix L Fennemann, Martijn Verdoes, and Olga Iliina. 2018. "Molecular Repolarisation of Tumour-Associated Macrophages." *Molecules (Basel, Switzerland)* 24 (1). <https://doi.org/10.3390/molecules24010009>.
- DiLeo, Matthew V., Gary D. Strahan, Meghan den Bakker, and Owen A. Hoekenga. 2011. "Weighted Correlation Network Analysis (WGCNA) Applied to the Tomato Fruit Metabolome." Edited by Peter Csermely. *PLoS ONE* 6 (10): e26683. <https://doi.org/10.1371/journal.pone.0026683>.
- Djebali, S, C A Davis, A Merkel, A Dobin, T Lassmann, A Mortazavi, A Tanzer, et al. 2012. "Landscape of Transcription in Human Cells" 489 (7414): 101–8. <https://doi.org/10.1038/nature11233>.
- DP Ryan, TS Hong, N Bardeesy. 2014. "Pancreatic Adenocarcinoma." *N. Engl. J. Med.* 371: 2140–41.
- Du, Zhou, Tong Sun, Ezgi Hacisuleyman, Teng Fei, Xiaodong Wang, Myles Brown, John L. Rinn, et al. 2016. "Integrative Analyses Reveal a Long Noncoding RNA-Mediated Sponge Regulatory Network in Prostate Cancer." *Nature Communications* 7 (1): 10982. <https://doi.org/10.1038/ncomms10982>.
- Dvinge, Heidi, Eunhee Kim, Omar Abdel-Wahab, and Robert K. Bradley. 2016. "RNA Splicing Factors as Oncoproteins and Tumour Suppressors." *Nature Reviews Cancer* 16 (7): 413–30. <https://doi.org/10.1038/nrc.2016.51>.
- Espy, M J, J R Uhl, L M Sloan, S P Buckwalter, M F Jones, E A Vetter, J D C Yao, et al. 2006. "Real-Time PCR in Clinical Microbiology: Applications for Routine Laboratory Testing." *Clinical Microbiology Reviews* 19 (1): 165–256. <https://doi.org/10.1128/CMR.19.1.165-256.2006>.
- Fabregat, Antonio, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, et al. 2018. "The Reactome Pathway Knowledgebase." *Nucleic Acids Research* 46 (D1): D649–55.

- <https://doi.org/10.1093/nar/gkx1132>.
- Feig, C., A. Gopinathan, A. Neesse, D. S. Chan, N. Cook, and D. A. Tuveson. 2012. "The Pancreas Cancer Microenvironment." *Clinical Cancer Research* 18 (16): 4266–76. <https://doi.org/10.1158/1078-0432.CCR-11-3114>.
- Frankish, Adam, Mark Diekhans, Anne-Maud Ferreira, Rory Johnson, Irwin Jungreis, Jane Loveland, Jonathan M Mudge, et al. 2019. "GENCODE Reference Annotation for the Human and Mouse Genomes." *Nucleic Acids Research* 47 (D1): D766–73. <https://doi.org/10.1093/nar/gky955>.
- Freedman, Andrew N., Carrie N. Klabunde, Kristine Wiant, Lindsey Enewold, Stacy W. Gray, Kelly K. Filipowski, Nancy L. Keating, et al. 2018. "Use of Next-Generation Sequencing Tests to Guide Cancer Treatment: Results From a Nationally Representative Survey of Oncologists in the United States." *JCO Precision Oncology*, no. 2 (November): 1–13. <https://doi.org/10.1200/PO.18.00169>.
- Friedenson, Bernard. 2005. "BRCA1 and BRCA2 Pathways and the Risk of Cancers Other than Breast or Ovarian." *MedGenMed: Medscape General Medicine* 7 (2): 60. <http://www.ncbi.nlm.nih.gov/pubmed/16369438>.
- Friedman, Adam A., Anthony Letai, David E. Fisher, and Keith T. Flaherty. 2015. "Precision Medicine for Cancer with Next-Generation Functional Diagnostics." *Nature Reviews Cancer* 15 (12): 747–56. <https://doi.org/10.1038/nrc4015>.
- Fu, Li, Yan-Ru Qin, Xiao-Yan Ming, Xian-Bo Zuo, Yu-Wen Diao, Li-Yi Zhang, Jiaoyu Ai, et al. 2017. "RNA Editing of SLC22A3 Drives Early Tumor Invasion and Metastasis in Familial Esophageal Cancer." *Proceedings of the National Academy of Sciences of the United States of America* 114 (23): E4631–40. <https://doi.org/10.1073/pnas.1703178114>.
- Fuller, Tova F., Anatole Ghazalpour, Jason E. Aten, Thomas A. Drake, Aldons J. Lusis, and Steve Horvath. 2007. "Weighted Gene Coexpression Network Analysis Strategies Applied to Mouse Weight." *Mammalian Genome* 18 (6–7): 463–72. <https://doi.org/10.1007/s00335-007-9043-3>.
- Galante, Pedro Alexandre Favoretto, Noboru Jo Sakabe, Natanja Kirschbaum-Slager, and Sandro José de Souza. 2004. "Detection and Evaluation of Intron Retention Events in the Human Transcriptome." *RNA (New York, N.Y.)* 10 (5): 757–65. <https://doi.org/10.1261/rna.5123504>.

- Gerken, Thomas A, Jayalakshmi Raman, Timothy A Fritz, and Oliver Jamison. 2006. "Identification of Common and Unique Peptide Substrate Preferences for the UDP-GalNAc:Polypeptide Alpha-N-Acetylgalactosaminyltransferases T1 and T2 Derived from Oriented Random Peptide Substrates." *The Journal of Biological Chemistry* 281 (43): 32403–16. <https://doi.org/10.1074/jbc.M605149200>.
- Gilbert, Walter. 1978. "Why Genes in Pieces?" *Nature* 271 (5645): 501–501. <https://doi.org/10.1038/271501a0>.
- Giulietti, Matteo, Alessandra Righetti, Giovanni Principato, and Francesco Piva. 2018. "LncRNA Co-Expression Network Analysis Reveals Novel Biomarkers for Pancreatic Cancer." *Carcinogenesis* 39 (8): 1016–25. <https://doi.org/10.1093/carcin/bgy069>.
- Golan, Talia, Pascal Hammel, Michele Reni, Eric Van Cutsem, Teresa Macarulla, Michael J. Hall, Joon-Oh Park, et al. 2019. "Maintenance Olaparib for Germline BRCA-Mutated Metastatic Pancreatic Cancer." *New England Journal of Medicine*, June, NEJMoa1903387. <https://doi.org/10.1056/NEJMoa1903387>.
- Golan, Talia, Hedy L. Kindler, Joon Oh Park, Michele Reni, Teresa Macarulla Mercade, Pascal Hammel, Eric Van Cutsem, et al. 2018. "Geographic and Ethnic Heterogeneity in the BRCA1/2 Pre-Screening Population for the Randomized Phase III POLO Study of Olaparib Maintenance in Metastatic Pancreatic Cancer (MPC)." *Journal of Clinical Oncology* 36 (15_suppl): 4115–4115. https://doi.org/10.1200/JCO.2018.36.15_suppl.4115.
- Gonzalez, Inma, Roberto Munita, Eneritz Agirre, Travis A Dittmer, Katia Gysling, Tom Misteli, and Reini F Luco. 2015. "A LncRNA Regulates Alternative Splicing via Establishment of a Splicing-Specific Chromatin Signature." *Nature Structural & Molecular Biology*. <https://doi.org/10.1038/nsmb.3005>.
- Grant, Robert C., Iris Selander, Ashton A. Connor, Shamini Selvarajah, Ayelet Borgida, Laurent Briollais, Gloria M. Petersen, Jordan Lerner-Ellis, Spring Holter, and Steven Gallinger. 2015. "Prevalence of Germline Mutations in Cancer Predisposition Genes in Patients With Pancreatic Cancer." *Gastroenterology* 148 (3): 556–64. <https://doi.org/10.1053/J.GASTRO.2014.11.042>.
- GUILLAUDEAU, Angélique, Karine DURAND, HÉLÈNE Rabinovitch-Chable,

- Isabelle POMMEPUY, LAURA Mesturoux, Sandrine ROBERT, ALAIN Chaunavel, Jean-Jacques MOREAU, and FRANÇOIS Labrousse. 2012. "Adult Diffuse Gliomas Produce MRNA Transcripts Encoding EGFR Isoforms Lacking a Tyrosine Kinase Domain." *International Journal of Oncology* 40 (4): 1142–52. <https://doi.org/10.3892/ijo.2011.1287>.
- Gupta, Rajnish A., Nilay Shah, Kevin C. Wang, Jeewon Kim, Hugo M. Horlings, David J. Wong, Miao-Chih Tsai, et al. 2010. "Long Non-Coding RNA HOTAIR Reprograms Chromatin State to Promote Cancer Metastasis." *Nature* 464 (7291): 1071–76. <https://doi.org/10.1038/nature08975>.
- Hajdu, Steven I. 2011a. "A Note from History: Landmarks in History of Cancer, Part 1." *Cancer* 117 (March): 1097–1102. <https://doi.org/10.1002/cncr.25553>.
- . 2011b. "A Note from History: Landmarks in History of Cancer, Part 2." *Cancer* 117 (12): 2811–20. <https://doi.org/10.1002/cncr.25825>.
- Hata, Shoji, Fujiko Kitamura, Midori Yamaguchi, Hiroshi Shitara, Makoto Murakami, and Hiroyuki Sorimachi. 2016. "A Gastrointestinal Calpain Complex, G-Calpain, Is a Heterodimer of CAPN8 and CAPN9 Calpain Isoforms, Which Play Catalytic and Regulatory Roles, Respectively." *The Journal of Biological Chemistry* 291 (53): 27313–22. <https://doi.org/10.1074/jbc.M116.763912>.
- Hayward, G.S., and M.G. Smith. 1972. "The Chromosome of Bacteriophage T5: I. Analysis of the Single-Stranded DNA Fragments by Agarose Gel Electrophoresis." *Journal of Molecular Biology* 63 (3): 383–95. [https://doi.org/10.1016/0022-2836\(72\)90435-4](https://doi.org/10.1016/0022-2836(72)90435-4).
- Heiland, Dieter Henrik, Irina Mader, Pascal Schlosser, Dietmar Pfeifer, Maria Stella Carro, Thomas Lange, Ralf Schwarzwald, Ioannis Vasilikos, Horst Urbach, and Astrid Weyerbrock. 2016. "Integrative Network-Based Analysis of Magnetic Resonance Spectroscopy and Genome Wide Expression in Glioblastoma Multiforme." *Scientific Reports* 6 (1): 29052. <https://doi.org/10.1038/srep29052>.
- Hodi, F. Stephen, Steven J. O'Day, David F. McDermott, Robert W. Weber, Jeffrey A. Sosman, John B. Haanen, Rene Gonzalez, et al. 2010. "Improved Survival with Ipilimumab in Patients with Metastatic Melanoma." *New England Journal of Medicine* 363 (8): 711–23.

- <https://doi.org/10.1056/NEJMoa1003466>.
- Hombach, Sonja, and Markus Kretz. 2016. "Non-Coding RNAs: Classification, Biology and Functioning." In *Advances in Experimental Medicine and Biology*, 937:3–17. https://doi.org/10.1007/978-3-319-42059-2_1.
- Hruban, RH. 2004. "An Illustrated Consensus on the Classification of Pancreatic Intraepithelial Neoplasia and Intraductal Papillary Mucinous Neoplasms." *Am. J. Surg. Pathol.* 28: 977–87.
- Hsu, Cheng-Ming, Pai-Mei Lin, Jan-Gowth Chang, Hsin-Ching Lin, Shau-Hsuan Li, Sheng-Fung Lin, and Ming-Yu Yang. 2017. "Upregulated SLC22A3 Has a Potential for Improving Survival of Patients with Head and Neck Squamous Cell Carcinoma Receiving Cisplatin Treatment." *Oncotarget* 8 (43): 74348–58. <https://doi.org/10.18632/oncotarget.20637>.
- Hu, Jiao-Jiao, Wei Song, Shao-Dan Zhang, Xiao-Hui Shen, Xue-Mei Qiu, Hua-Zhang Wu, Pi-Hai Gong, et al. 2016. "HBx-Upregulated LncRNA UCA1 Promotes Cell Growth and Tumorigenesis by Recruiting EZH2 and Repressing P27Kip1/CDK2 Signaling." *Scientific Reports* 6 (1): 23521. <https://doi.org/10.1038/srep23521>.
- Huarte, Maite. 2015. "The Emerging Role of LncRNAs in Cancer." *Nature Medicine* 21 (11): 1253–61. <https://doi.org/10.1038/nm.3981>.
- Huarte, Maite, and John L. Rinn. 2010. "Large Non-Coding RNAs: Missing Links in Cancer?" *Human Molecular Genetics* 19 (R2): R152–61. <https://doi.org/10.1093/hmg/ddq353>.
- Hudson (Chairperson), Thomas J., Warwick Anderson, Axel Aretz, Anna D. Barker, Cindy Bell, Rosa R. Bernabé, M. K. Bhan, et al. 2010. "International Network of Cancer Genome Projects." *Nature* 464 (7291): 993–98. <https://doi.org/10.1038/nature08987>.
- Iacobuzio-Donahue, Christine A., Anirban Maitra, Grace L. Shen-Ong, Tjarda van Heek, Raheela Ashfaq, Renee Meyer, Kimberly Walter, et al. 2002. "Discovery of Novel Tumor Markers of Pancreatic Cancer Using Global Gene Expression Technology." *The American Journal of Pathology* 160 (4): 1239–49. [https://doi.org/10.1016/S0002-9440\(10\)62551-5](https://doi.org/10.1016/S0002-9440(10)62551-5).
- INCA. 2018. "Câncer de Pâncreas." 2018. <https://www.inca.gov.br/tipos-de-cancer/cancer-de-pancreas>.
- Iyer, Matthew K, Yashar S Niknafs, Rohit Malik, Udit Singhal, Anirban Sahu,

- Yasuyuki Hosono, Terrence R Barrette, et al. 2015. "The Landscape of Long Noncoding RNAs in the Human Transcriptome." *Nature Genetics* 47 (3): 199–208. <https://doi.org/10.1038/ng.3192>.
- Jeong, Seonhyang, Jandee Lee, Daham Kim, Mi-Youn Seol, Woo Kyung Lee, Jong Ju Jeong, Kee-Hyun Nam, et al. 2016. "Relationship of Focally Amplified Long Noncoding on Chromosome 1 (FAL1) LncRNA with E2F Transcription Factors in Thyroid Cancer." *Medicine* 95 (4): e2592. <https://doi.org/10.1097/MD.0000000000002592>.
- Jiang, Wei, Yaqing Zhang, Kevin T. Kane, Meredith A. Collins, Diane M. Simeone, Marina Pasca di Magliano, and Kevin Tri Nguyen. 2015. "CD44 Regulates Pancreatic Cancer Invasion through MT1-MMP." *Molecular Cancer Research* 13 (1): 9–15. <https://doi.org/10.1158/1541-7786.MCR-14-0076>.
- Jonckheere, Nicolas, Nicolas Skrypek, and Isabelle van Seuning. 2010. "Mucins and Pancreatic Cancer." *Cancers* 2 (4): 1794–1812. <https://doi.org/10.3390/cancers2041794>.
- Kanehisa, M., and Susumu Goto. 2000. "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 28 (1): 27–30. <https://doi.org/10.1093/nar/28.1.27>.
- Karolchik, Donna, Galt P. Barber, Jonathan Casper, Hiram Clawson, Melissa S. Cline, Mark Diekhans, Timothy R. Dreszer, et al. 2014. "The UCSC Genome Browser Database: 2014 Update." *Nucleic Acids Research* 42 (D1): D764–70. <https://doi.org/10.1093/nar/gkt1168>.
- Kassambara, Alboukadel, Marcin Kosinski, Przemyslaw Biecek, and Scheipl Fabian. 2019. "Package 'survminer' Type Package Title Drawing Survival Curves Using 'Ggplot2.'" <https://cran.r-project.org/web/packages/survminer/survminer.pdf>.
- Katz, Yarden, Eric T. Wang, Jacob Silterra, Schraga Schwartz, Bang Wong, Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov, Edoardo M. Airoldi, and Christopher B. Burge. 2015. "Quantitative Visualization of Alternative Exon Expression from RNA-Seq Data." *Bioinformatics* 31 (14): 2400–2402. <https://doi.org/10.1093/bioinformatics/btv034>.
- Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and a. D. Haussler. 2002. "The Human Genome Browser at UCSC." *Genome*

- Research* 12 (6): 996–1006. <https://doi.org/10.1101/gr.229102>.
- Khurshid, Safiya. 2019. “Abstract 2639: Insulin Receptor Alternative Splicing during Development and Cancer.” In *Molecular and Cellular Biology / Genetics*, 79:2639–2639. American Association for Cancer Research. <https://doi.org/10.1158/1538-7445.SABCS18-2639>.
- Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L Salzberg. 2013. “TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions.” *Genome Biology* 14 (4): R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
- Kim, Eddo, Alon Magen, and Gil Ast. 2007. “Different Levels of Alternative Splicing among Eukaryotes.” *Nucleic Acids Research* 35 (1): 125–31. <https://doi.org/10.1093/nar/gkl924>.
- Kleihues, Paul, and Leslie H. Sobin. 2000. “World Health Organization Classification of Tumors.” *Cancer* 88 (12): 2887–2887. [https://doi.org/10.1002/1097-0142\(20000615\)88:12<2887::AID-CNCR32>3.0.CO;2-F](https://doi.org/10.1002/1097-0142(20000615)88:12<2887::AID-CNCR32>3.0.CO;2-F).
- Klein, Carl H. Von. 1905. “THE MEDICAL FEATURES OF THE PAPYRUS EBERS.” *JAMA: The Journal of the American Medical Association* XLV (26): 1928. <https://doi.org/10.1001/jama.1905.52510260014001e>.
- Kleinhans, Regina, Martin Woywod, Marieluise Schmidbauer, Walter Wirths, and Eike Kottkamp. 2018. “Application Note: Cell Growth Analysis Cell Growth Analysis Using the 24-Channel Microscope ZenCELL Owl.” www.zencellowl.com.
- Kodzius, Rimantas, Miki Kojima, Hiromi Nishiyori, Mari Nakamura, Shiro Fukuda, Michihira Tagami, Daisuke Sasaki, et al. 2006. “CAGE: Cap Analysis of Gene Expression.” *Nature Methods* 3 (3): 211–22. <https://doi.org/10.1038/nmeth0306-211>.
- Kogo, R., T. Shimamura, K. Mimori, K. Kawahara, S. Imoto, T. Sudo, F. Tanaka, et al. 2011. “Long Noncoding RNA HOTAIR Regulates Polycomb-Dependent Chromatin Modification and Is Associated with Poor Prognosis in Colorectal Cancers.” *Cancer Research* 71 (20): 6320–26. <https://doi.org/10.1158/0008-5472.CAN-11-1021>.
- Koikawa, Kazuhiro, Kenoki Ohuchida, Shin Takesue, Yohei Ando, Shin Kibe, Hiromichi Nakayama, Sho Endo, et al. 2018. “Pancreatic Stellate Cells

- Reorganize Matrix Components and Lead Pancreatic Cancer Invasion via the Function of Endo180.” *Cancer Letters* 412 (January): 143–54. <https://doi.org/10.1016/j.canlet.2017.10.010>.
- Kong, Jianlu, Wenjie Sun, Chen Li, Ledong Wan, Shuo Wang, Yihua Wu, Enping Xu, Honghe Zhang, and Maode Lai. 2016. “Long Non-Coding RNA LINC01133 Inhibits Epithelial–Mesenchymal Transition and Metastasis in Colorectal Cancer B1. Kong J, Sun W, Li C, Wan L, Wang S, Wu Y, et Al. Long Non-Coding RNA LINC01133 Inhibits Epithelial–Mesenchymal Transition and Metastasis in Co.” *Cancer Letters* 380 (2): 476–84. <https://doi.org/10.1016/j.canlet.2016.07.015>.
- Kushi, Lawrence H., Colleen Doyle, Marji McCullough, Cheryl L. Rock, Wendy Demark-Wahnefried, Elisa V. Bandera, Susan Gapstur, Alpa V. Patel, Kimberly Andrews, and Ted Gansler. 2012. “American Cancer Society Guidelines on Nutrition and Physical Activity for Cancer Prevention.” *CA: A Cancer Journal for Clinicians* 62 (1): 30–67. <https://doi.org/10.3322/caac.20140>.
- Lamere, Alicia T., and Jun Li. 2019. “Inference of Gene Co-Expression Networks from Single-Cell RNA-Sequencing Data.” In , 141–53. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-9057-3_10.
- Lander, Eric S., Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–921. <https://doi.org/10.1038/35057062>.
- Langfelder, Peter, and Steve Horvath. 2008. “WGCNA: An R Package for Weighted Correlation Network Analysis.” *BMC Bioinformatics* 9 (1): 559. <https://doi.org/10.1186/1471-2105-9-559>.
- Lee, Genee Y, Paraic A Kenny, Eva H Lee, and Mina J Bissell. 2007. “Three-Dimensional Culture Models of Normal and Malignant Breast Epithelial Cells.” *Nature Methods* 4 (4): 359–65. <https://doi.org/10.1038/nmeth1015>.
- Lefever, Steve, Jasper Anckaert, Pieter-Jan Volders, Manuel Luybaert, Jo Vandesompele, and Pieter Mestdagh. 2017. “DecodeRNA- Predicting Non-Coding RNA Functions Using Guilt-by-Association.” *Database: The Journal of Biological Databases and Curation* 2017. <https://doi.org/10.1093/database/bax042>.

- Li, Bo, and Colin N Dewey. 2011. "RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome." *BMC Bioinformatics* 12 (1): 323. <https://doi.org/10.1186/1471-2105-12-323>.
- Li, Hui-Jin, Xu Li, Huan Pang, Jing-Jing Pan, Xiao-Juan Xie, and Wei Chen. 2015. "Long Non-Coding RNA UCA1 Promotes Glutamine Metabolism by Targeting MiR-16 in Human Bladder Cancer." *Japanese Journal of Clinical Oncology* 45 (11): 1055–63. <https://doi.org/10.1093/jjco/hyv132>.
- Li, Jiao, Dejun Liu, Rong Hua, Junfeng Zhang, Wei Liu, Yanmiao Huo, Yifan Cheng, Jie Hong, and Yongwei Sun. 2014. "Long Non-Coding RNAs Expressed in Pancreatic Ductal Adenocarcinoma and LncRNA BC008363 an Independent Prognostic Factor in PDAC." *Pancreatology* 14 (5): 385–90. <https://doi.org/10.1016/j.pan.2014.07.013>.
- Li, Jiarui, Zhe Li, Haoxuan Kan, Zhao Sun, Jiazhang Xing, Yuejuan Cheng, and Chunmei Bai. 2019. "CA19-9 Elevation as an Indication to Start Salvage Treatment in Surveillance after Pancreatic Cancer Resection." *Pancreatology* 19 (2): 302–6. <https://doi.org/10.1016/J.PAN.2019.01.023>.
- Li, R-K, J- Gao, L-H Guo, G-Q Huang, and W-H Luo. 2017. "PTENP1 Acts as a CeRNA to Regulate PTEN by Sponging MiR-19b and Explores the Biological Role of PTENP1 in Breast Cancer." *Cancer Gene Therapy* 24 (7): 309–15. <https://doi.org/10.1038/cgt.2017.29>.
- Li, Wan, Han Jinhua, Liu Ting, Dong Shunli, Xie Feng, Chen Hongxia, and Huang Jun. 2013. "Scaffolding Protein SPIDR/KIAA0146 Connects the Bloom Syndrome Helicase with Homologous Recombination Repair." *PNAS* 110 (26): 10646–51. <https://doi.org/10.1073/pnas.1220921110>.
- Li, Zhengkun, Xu Li, Shouzhen Wu, Mei Xue, and Wei Chen. 2014. "Long Non-Coding RNA UCA1 Promotes Glycolysis by Upregulating Hexokinase 2 through the MTOR-STAT3/MicroRNA143 Pathway." *Cancer Science* 105 (8): 951–55. <https://doi.org/10.1111/cas.12461>.
- Li, Zhihua, Xiaohui Zhao, Yu Zhou, Yimin Liu, Quanbo Zhou, Huilin Ye, YinXue Wang, et al. 2015. "The Long Non-Coding RNA HOTTIP Promotes Progression and Gemcitabine Resistance by Regulating HOXA13 in Pancreatic Cancer." *Journal of Translational Medicine* 13 (1): 84. <https://doi.org/10.1186/s12967-015-0442-z>.
- Liu, Ai-Na, Hua-Jun Qu, Cai-Yan Yu, and Ping Sun. 2018. "Knockdown of

- LINC01614 Inhibits Lung Adenocarcinoma Cell Progression by Up-regulating MiR-217 and Down-regulating *FOXP1*.” *Journal of Cellular and Molecular Medicine* 22 (9): 4034–44. <https://doi.org/10.1111/jcmm.13483>.
- Liu, Hui, Junyun Luo, Siyu Luan, Chongsheng He, and Zhaoyong Li. 2019. “Long Non-Coding RNAs Involved in Cancer Metabolic Reprogramming.” *Cellular and Molecular Life Sciences* 76 (3): 495–504. <https://doi.org/10.1007/s00018-018-2946-1>.
- Livak, K J, and T D Schmittgen. 2001. “Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR And.” *Methods* 25: 402–8. <https://doi.org/10.1006/meth.2001.1262>.
- Locker, Gershon Y., Stanley Hamilton, Jules Harris, John M. Jessup, Nancy Kemeny, John S. Macdonald, Mark R. Somerfield, Daniel F. Hayes, and Robert C. Bast. 2006. “ASCO 2006 Update of Recommendations for the Use of Tumor Markers in Gastrointestinal Cancer.” *Journal of Clinical Oncology* 24 (33): 5313–27. <https://doi.org/10.1200/JCO.2006.08.2644>.
- Lopez-Pajares, Vanessa, Kun Qu, Jiajing Zhang, Dan E. Webster, Brook C. Barajas, Zurab Siprashvili, Brian J. Zarnegar, et al. 2015. “A LncRNA-MAF:MAFB Transcription Factor Network Regulates Epidermal Differentiation.” *Developmental Cell* 32 (6): 693–706. <https://doi.org/10.1016/J.DEVCEL.2015.01.028>.
- Lynch, Thomas J., Daphne W. Bell, Raffaella Sordella, Sarada Gurubhagavatula, Ross A. Okimoto, Brian W. Brannigan, Patricia L. Harris, et al. 2004. “Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib.” *New England Journal of Medicine* 350 (21): 2129–39. <https://doi.org/10.1056/NEJMoa040938>.
- Mackay, John F, Christopher D Wright, and Roderick G Bonfiglioli. 2008. “A New Approach to Varietal Identification in Plants by Microsatellite High Resolution Melting Analysis: Application to the Verification of Grapevine and Olive Cultivars.” *Plant Methods* 4 (1): 8. <https://doi.org/10.1186/1746-4811-4-8>.
- Magalhães, Yuli T., Jessica O. Farias, Lucas F. Monteiro, and Fábio L. Forti. 2018. “Measuring the Contributions of the Rho Pathway to the DNA Damage Response in Tumor Epithelial Cells.” In *Methods in Molecular Biology (Clifton, N.J.)*, 1821:339–55. https://doi.org/10.1007/978-1-4939-8612-5_23.

- Maher, Christopher A., Chandan Kumar-Sinha, Xuhong Cao, Shanker Kalyana-Sundaram, Bo Han, Xiaojun Jing, Lee Sam, Terrence Barrette, Nallasivam Palanisamy, and Arul M. Chinnaiyan. 2009. "Transcriptome Sequencing to Detect Gene Fusions in Cancer." *Nature* 458 (7234): 97–101. <https://doi.org/10.1038/nature07638>.
- Mann, D V, R Edwards, S Ho, W Y Lau, and G Glazer. 2000. "Elevated Tumour Marker CA19-9: Clinical Interpretation and Influence of Obstructive Jaundice." *European Journal of Surgical Oncology: The Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* 26 (5): 474–79. <https://doi.org/10.1053/ejso.1999.0925>.
- Marguerat, Samuel, and Jürg Bähler. 2010. "RNA-Seq: From Technology to Biology." *Cellular and Molecular Life Sciences* 67 (4): 569–79. <https://doi.org/10.1007/s00018-009-0180-6>.
- Marinov, Georgi K., Brian A. Williams, Ken McCue, Gary P. Schroth, Jason Gertz, Richard M. Myers, and Barbara J. Wold. 2014. "From Single-Cell to Cell-Pool Transcriptomes: Stochasticity in Gene Expression and RNA Splicing." *Genome Research* 24 (3). <https://doi.org/10.1101/gr.161034.113>.
- Martin, Lance, and Howard Y Chang. 2012. "Uncovering the Role of Genomic "Dark Matter" in Human Disease." *The Journal of Clinical Investigation* 122 (5): 1589–95. <https://doi.org/10.1172/JCI60020>.
- Maurer, Carlo, Sam R Holmstrom, Jing He, Pasquale Laise, Tao Su, Aqeel Ahmed, Hanina Hibshoosh, et al. 2019. "Experimental Microdissection Enables Functional Harmonisation of Pancreatic Cancer Subtypes." *Gut* 68 (6): 1034–43. <https://doi.org/10.1136/gutjnl-2018-317706>.
- Mekapogu, Alpha R., Srinivasa P. Pothula, Romano C. Pirola, Jeremy S. Wilson, and Minoti V. Apte. 2019. "Multifunctional Role of Pancreatic Stellate Cells in Pancreatic Cancer." *Annals of Pancreatic Cancer* 2 (0): 10–10. <https://doi.org/10.21037/apc.2019.05.02>.
- Mironov, A A, J W Fickett, and M S Gelfand. 1999. "Frequent Alternative Splicing of Human Genes." *Genome Research* 9 (12): 1288–93. <https://doi.org/10.1101/gr.9.12.1288>.
- Moffitt, Richard A, Raoud Marayati, Elizabeth L Flate, Keith E Volmar, S Gabriela Herrera Loeza, Katherine A Hoadley, Naim U Rashid, et al. 2015. "Virtual Microdissection Identifies Distinct Tumor- and Stroma-Specific Subtypes of

- Pancreatic Ductal Adenocarcinoma.” *Nature Genetics* 47 (10): 1168–78. <https://doi.org/10.1038/ng.3398>.
- Moraes, Luís. 2019. “Desenvolvimento de Xenotransplantes de Tumores Pancreáticos Humanos Para Varredura Genética de Alvos Moleculares Com Potencial Terapêutico.” São Paulo: Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo. <https://doi.org/10.11606/T.46.2019.tde-07052019-110949>.
- Morlando, Mariangela, Monica Ballarino, and Alessandro Fatica. 2015. “Long Non-Coding RNAs: New Players in Hematopoiesis and Leukemia.” *Frontiers in Medicine* 2: 23. <https://doi.org/10.3389/fmed.2015.00023>.
- Mukherjee, Shradha. 2019. “Genomics-Guided Immunotherapy for Precision Medicine in Cancer.” *Cancer Biotherapy and Radiopharmaceuticals*, July, cbr.2018.2758. <https://doi.org/10.1089/cbr.2018.2758>.
- Müller, Sören, Susanne Raulefs, Philipp Bruns, Fabian Afonso-Grunz, Anne Plötner, Rolf Thermann, Carsten Jäger, et al. 2015. “Next-Generation Sequencing Reveals Novel Differentially Regulated MRNAs, LncRNAs, MiRNAs, SdRNAs and a PiRNA in Pancreatic Cancer.” *Molecular Cancer* 14 (1): 94. <https://doi.org/10.1186/s12943-015-0358-5>.
- Neesse, Albrecht, Christian Alexander Bauer, Daniel Öhlund, Matthias Lauth, Malte Buchholz, Patrick Michl, David A Tuveson, and Thomas M Gress. 2019. “Stromal Biology and Therapy in Pancreatic Cancer: Ready for Clinical Translation?” *Gut* 68 (1): 159–71. <https://doi.org/10.1136/gutjnl-2018-316451>.
- Neoptolemos, John P., Jörg Kleeff, Patrick Michl, Eithne Costello, William Greenhalf, and Daniel H. Palmer. 2018. “Therapeutic Developments in Pancreatic Cancer: Current and Future Perspectives.” *Nature Reviews Gastroenterology & Hepatology* 15 (6): 333–48. <https://doi.org/10.1038/s41575-018-0005-x>.
- Nguyen, Andrew H., Irmina A. Elliott, Nanping Wu, Cynthia Matsumura, Maria Vogelauer, Narsis Attar, Amanda Dann, et al. 2017. “Histone Deacetylase Inhibitors Provoke a Tumor Supportive Phenotype in Pancreatic Cancer Associated Fibroblasts.” *Oncotarget* 8 (12): 19074–88. <https://doi.org/10.18632/oncotarget.13572>.
- Notta, Faiyaz, Michelle Chan-Seng-Yue, Mathieu Lemire, Yilong Li, Gavin W.

- Wilson, Ashton A. Connor, Robert E. Denroche, et al. 2016. "A Renewed Model of Pancreatic Cancer Evolution Based on Genomic Rearrangement Patterns." *Nature* 538 (7625): 378–82. <https://doi.org/10.1038/nature19823>.
- Nyga, Agata, Umber Cheema, and Marilena Loizidou. 2011. "3D Tumour Models: Novel in Vitro Approaches to Cancer Studies." *Journal of Cell Communication and Signaling* 5 (3): 239–48. <https://doi.org/10.1007/s12079-011-0132-4>.
- Oltean, S, and D O Bates. 2014. "Hallmarks of Alternative Splicing in Cancer." *Oncogene* 33 (46): 5311–18. <https://doi.org/10.1038/onc.2013.533>.
- Padfield, Emily, Hayley P. Ellis, and Kathreena M. Kurian. 2015. "Current Therapeutic Advances Targeting EGFR and EGFRvIII in Glioblastoma." *Frontiers in Oncology* 5 (January): 5. <https://doi.org/10.3389/fonc.2015.00005>.
- Parasramka, Mansi A., Sayantan Maji, Akiko Matsuda, Irene K. Yan, and Tushar Patel. 2016. "Long Non-Coding RNAs as Novel Targets for Therapy in Hepatocellular Carcinoma." *Pharmacology & Therapeutics* 161 (May): 67–78. <https://doi.org/10.1016/j.pharmthera.2016.03.004>.
- Patro, Rob, Geet Duggal, Michael I Love, Rafael A Irizarry, and Carl Kingsford. 2017. "Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression." *Nature Methods* 14 (4): 417–19. <https://doi.org/10.1038/nmeth.4197>.
- Pertea, Mihaela, Geo M Pertea, Corina M Antonescu, Tsung-Cheng Chang, Joshua T Mendell, and Steven L Salzberg. 2015. "StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads." *Nature Biotechnology* 33 (3): 290–95. <https://doi.org/10.1038/nbt.3122>.
- Pickl, M, and C H Ries. 2009. "Comparison of 3D and 2D Tumor Models Reveals Enhanced HER2 Activation in 3D Associated with an Increased Response to Trastuzumab." *Oncogene* 28 (3): 461–68. <https://doi.org/10.1038/onc.2008.394>.
- Poruk, Katherine E, D Z Gay, K Brown, J D Mulvihill, K M Boucher, C L Scaife, M A Firpo, and S J Mulvihill. 2013. "The Clinical Utility of CA 19-9 in Pancreatic Adenocarcinoma: Diagnostic and Prognostic Updates." *Current Molecular Medicine* 13 (3): 340–51. <https://doi.org/10.2174/1566524011313030003>.
- Puleo, Francesco, Rémy Nicolle, Yuna Blum, Jérôme Cros, Laetitia Marisa,

- Pieter Demetter, Eric Quertinmont, et al. 2018. "Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features." *Gastroenterology* 155 (6): 1999-2013.e3. <https://doi.org/10.1053/j.gastro.2018.08.033>.
- Quagliata, Luca, Matthias S. Matter, Salvatore Piscuoglio, Leila Arabi, Christian Ruiz, Alfredo Procino, Michal Kovac, et al. 2014. "Long Noncoding RNA HOTTIP/HOXA13 Expression Is Associated with Disease Progression and Predicts Outcome in Hepatocellular Carcinoma Patients." *Hepatology* 59 (3): 911–23. <https://doi.org/10.1002/hep.26740>.
- Rahib, L. 2014. "Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States." *Cancer Res.* 74: 2913–21.
- Rahib, Lola, Benjamin D. Smith, Rhonda Aizenberg, Allison B. Rosenzweig, Julie M. Fleshman, and Lynn M. Matrisian. 2014. "Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States." *Cancer Research* 74 (11): 2913–21. <https://doi.org/10.1158/0008-5472.CAN-14-0155>.
- Raphael, Benjamin J. B.J., Ralph H. RH. Hruban, A.J. Andrew J. Aguirre, RA. Richard A. Moffitt, J.J. Jen Jen Yeh, Chip Stewart, A. Gordon Robertson, et al. 2017. "Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma." *Cancer Cell* 32 (2): 185-203.e13. <https://www.sciencedirect.com/science/article/pii/S1535610817302994>.
- Raphael, Benjamin J, Ralph H Hruban, Andrew J Aguirre, Richard A Moffitt, Jen Jen Yeh, Chip Stewart, A Gordon Robertson, et al. 2017. "Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma." *Cancer Cell* 32: 185-203.e13. <https://doi.org/10.1016/j.ccell.2017.07.007>.
- Rathinasamy, Baskaran, and Bharath Kumar Velmurugan. 2018. "Role of LncRNAs in the Cancer Development and Progression and Their Regulation by Various Phytochemicals." *Biomedicine & Pharmacotherapy* 102 (June): 242–48. <https://doi.org/10.1016/J.BIOPHA.2018.03.077>.
- Reimand, Jüri, Tambet Arak, Priit Adler, Liis Kolberg, Sulev Reisberg, Hedi Peterson, and Jaak Vilo. 2016. "G:Profiler-a Web Server for Functional Interpretation of Gene Lists (2016 Update)." *Nucleic Acids Research* 44 (W1): W83-9. <https://doi.org/10.1093/nar/gkw199>.

- Reis, Eduardo M., and Sergio Verjovski-Almeida. 2012a. "Perspectives of Long Non-Coding RNAs in Cancer Diagnostics." *Frontiers in Genetics* 3 (MAR): 1–8. <https://doi.org/10.3389/fgene.2012.00032>.
- . 2012b. "Perspectives of Long Non-Coding RNAs in Cancer Diagnostics." *Frontiers in Genetics* 3 (March): 32. <https://doi.org/10.3389/fgene.2012.00032>.
- RL Siegel, KD Miller, A Jemal. 2017. "Cancer Statistics, 2017." *CA Cancer J. Clin.* 67: 7–30.
- Sakabe, Noboru, and Sandro de Souza. 2007. "Sequence Features Responsible for Intron Retention in Human." *BMC Genomics* 8 (1): 59. <https://doi.org/10.1186/1471-2164-8-59>.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Ed. 2. Cold Spring Harbor Laboratory Press. <https://www.cabdirect.org/cabdirect/abstract/19901616061>.
- Santucci, Nicolas, Olivier Facy, Pablo Ortega-Deballon, Jean-Baptiste Lequeu, Paul Rat, and Patrick Rat. 2018. "CA 19–9 Predicts Resectability of Pancreatic Cancer Even in Jaundiced Patients." *Pancreatology* 18 (6): 666–70. <https://doi.org/10.1016/J.PAN.2018.07.001>.
- Scarpa, Aldo, William Mantovani, Paola Capelli, Stefania Beghelli, Letizia Boninsegna, Rossella Bettini, Francesco Panzuto, Paolo Pederzoli, Gianfranco delle Fave, and Massimo Falconi. 2010. "Pancreatic Endocrine Tumors: Improved TNM Staging and Histopathological Grading Permit a Clinically Efficient Prognostic Stratification of Patients." *Modern Pathology* 23 (6): 824–33. <https://doi.org/10.1038/modpathol.2010.58>.
- Schaukowitch, K., and T.-K. Kim. 2014. "Emerging Epigenetic Mechanisms of Long Non-Coding RNAs." *Neuroscience* 264 (April): 25–38. <https://doi.org/10.1016/J.NEUROSCIENCE.2013.12.009>.
- Schaukowitch, Katie, Jae-Yeol Joo, Xihui Liu, Jonathan K. Watts, Carlos Martinez, and Tae-Kyung Kim. 2014. "Enhancer RNA Facilitates NELF Release from Immediate Early Genes." *Molecular Cell* 56 (1): 29–42. <https://doi.org/10.1016/J.MOLCEL.2014.08.023>.
- Schmittgen, Thomas D, and Kenneth J Livak. 2008. "Analyzing Real-Time PCR Data by the Comparative C(T) Method." *Nature Protocols* 3 (6): 1101–8.

<http://www.ncbi.nlm.nih.gov/pubmed/18546601>.

- Schneider, Caroline A, Wayne S Rasband, and Kevin W Eliceiri. 2012. "NIH Image to ImageJ: 25 Years of Image Analysis." *Nature Methods* 9 (7): 671–75. <https://doi.org/10.1038/nmeth.2089>.
- Schwarzer, Adrian, Stephan Emmrich, Franziska Schmidt, Dominik Beck, Michelle Ng, Christina Reimer, Felix Ferdinand Adams, et al. 2017. "The Non-Coding RNA Landscape of Human Hematopoiesis and Leukemia." *Nature Communications* 8 (1): 218. <https://doi.org/10.1038/s41467-017-00212-4>.
- Scotti, Marina M., and Maurice S. Swanson. 2016. "RNA Mis-Splicing in Disease." *Nature Reviews Genetics* 17 (1): 19–32. <https://doi.org/10.1038/nrg.2015.3>.
- Seo, Y D, and V G Pillarisetty. 2017. "T-Cell Programming in Pancreatic Adenocarcinoma: A Review." *Cancer Gene Therapy* 24 (3): 106–13. <https://doi.org/10.1038/cgt.2016.66>.
- Serghiou, Stylianos, Aikaterini Kyriakopoulou, John P. A. Ioannidis, RP Alexander, G Fang, J Rozowsky, M Snyder, et al. 2016. "Long Noncoding RNAs as Novel Predictors of Survival in Human Cancer: A Systematic Review and Meta-Analysis." *Molecular Cancer* 15 (1): 50. <https://doi.org/10.1186/s12943-016-0535-1>.
- Seufferlein, Thomas, and Julia Mayerle. 2016. "Pancreatic Cancer in 2015: Precision Medicine in Pancreatic Cancer — Fact or Fiction?" *Nature Reviews Gastroenterology & Hepatology*, 2–3. <https://doi.org/10.1038/nrgastro.2015.215>.
- Shafee, Thomas, Rohan Lowe, and R Lowe. 2017. "Eukaryotic and Prokaryotic Gene Structure." *WikiJournal of Medicine* 4 (1): 2. <https://doi.org/10.15347/wjm/2017.002>.
- Shan, Yujia, Jia Ma, Yue Pan, Jialei Hu, Bing Liu, and Li Jia. 2018. "LncRNA SNHG7 Sponges MiR-216b to Promote Proliferation and Liver Metastasis of Colorectal Cancer through Upregulating GALNT1." *Cell Death & Disease* 9 (7): 722. <https://doi.org/10.1038/s41419-018-0759-7>.
- Shi, Xiaoshun, Haoming Tan, Xiaobing Le, Haibing Xian, Xiaoxiang Li, Kailing Huang, Viola Yingjun Luo, et al. 2018. "An Expression Signature Model to Predict Lung Adenocarcinoma-Specific Survival." *Cancer Management and Research* 10: 3717–32. <https://doi.org/10.2147/CMAR.S159563>.

- Shi, Yu, Weina Gao, Nikki K. Lytle, Peiwu Huang, Xiao Yuan, Amanda M. Dann, Maya Ridinger-Saison, et al. 2019. "Targeting LIF-Mediated Paracrine Interaction for Pancreatic Cancer Therapy and Monitoring." *Nature* 569 (7754): 131–35. <https://doi.org/10.1038/s41586-019-1130-6>.
- Shiraki, Toshiyuki, Shinji Kondo, Shintaro Katayama, Kazunori Waki, Takeya Kasukawa, Hideya Kawaji, Rimantas Kodzius, et al. 2003. "Cap Analysis Gene Expression for High-Throughput Analysis of Transcriptional Starting Point and Identification of Promoter Usage." *Proceedings of the National Academy of Sciences of the United States of America* 100 (26): 15776–81. <https://doi.org/10.1073/pnas.2136655100>.
- Siegel, Rebecca L., Kimberly D. Miller, and Ahmedin Jemal. 2016. "Cancer Statistics, 2016." *CA: A Cancer Journal for Clinicians* 66 (1): 7–30. <https://doi.org/10.3322/caac.21332>.
- Singh, Shiv, Marie Christin Hasselluhn, and Albrecht Neesse. 2019. "A Tangled Tale of Molecular Subtypes in Pancreatic Cancer" 68 (6). <https://doi.org/10.1136/gutjnl-2018-318086>.
- Siriwardena, A. K., and A. M. Siriwardena. 2014. "Pancreatic Cancer." *BMJ* 349 (oct31 2): g6385–g6385. <https://doi.org/10.1136/bmj.g6385>.
- Slamon, Dennis J., Brian Leyland-Jones, Steven Shak, Hank Fuchs, Virginia Paton, Alex Bajamonde, Thomas Fleming, et al. 2001. "Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2." *New England Journal of Medicine* 344 (11): 783–92. <https://doi.org/10.1056/NEJM200103153441101>.
- Slesak, Barbara, Antonina Harlozinska-Szmyrka, Witold Knast, Pawel Sedlaczek, Arie van Dalen, and Roland Einarsson. 2000. "Tissue Polypeptide Specific Antigen (TPS), a Marker for Differentiation between Pancreatic Carcinoma and Chronic Pancreatitis." *Cancer* 89 (1): 83–88. [https://doi.org/10.1002/1097-0142\(20000701\)89:1<83::AID-CNCR12>3.0.CO;2-J](https://doi.org/10.1002/1097-0142(20000701)89:1<83::AID-CNCR12>3.0.CO;2-J).
- Smith, Christof C., Sara R. Selitsky, Shengjie Chai, Paul M. Armistead, Benjamin G. Vincent, and Jonathan S. Serody. 2019. "Alternative Tumour-Specific Antigens." *Nature Reviews Cancer*, July, 1. <https://doi.org/10.1038/s41568-019-0162-4>.

- Soldan, Mônica. 2017. "Rastreamento Do Câncer de Pâncreas Pancreatic Cancer Screening." *Rev. Col. Bras. Cir* 44 (2): 109–11. <https://doi.org/10.1590/0100-69912017002015>.
- Song, Jukun, Qiuyan Xu, Haodeng Zhang, Xinhai Yin, Chen Zhu, Ke Zhao, and Jianguo Zhu. 2018. "Five Key LncRNAs Considered as Prognostic Targets for Predicting Pancreatic Ductal Adenocarcinoma." *Journal of Cellular Biochemistry* 119 (6): 4559–69. <https://doi.org/10.1002/jcb.26598>.
- Song, Yidan, Yihua Pan, and Jun Liu. 2019. "Functional Analysis of LncRNAs Based on Competitive Endogenous RNA in Tongue Squamous Cell Carcinoma." *PeerJ* 7 (May): e6991. <https://doi.org/10.7717/peerj.6991>.
- Sosa, Omar Julio. 2019a. "Identificação e Anotação Funcional de Novos Transcritos Com Expressão Alterada No Câncer Pancreático." São Paulo: Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo. <https://doi.org/10.11606/T.95.2019.tde-26042019-164116>.
- . 2019b. "Identificação e Anotação Funcional de Novos Transcritos Com Expressão Alterada No Câncer Pancreático." São Paulo: Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo. <https://doi.org/10.11606/T.95.2019.tde-26042019-164116>.
- Stathis, Anastasios, and Malcolm J. Moore. 2010. "Advanced Pancreatic Carcinoma: Current Treatment and Future Challenges." *Nature Reviews Clinical Oncology* 7 (3): 163–72. <https://doi.org/10.1038/nrclinonc.2009.236>.
- Sun, Miao, Shrikanth S Gadad, Dae-Seok Kim, and W Lee Kraus. 2015. "Discovery, Annotation, and Functional Analysis of Long Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells." *Molecular Cell* 59 (4): 698–711. <https://doi.org/10.1016/j.molcel.2015.06.023>.
- Sun, Qiqing, Bo Zhang, Qiangsheng Hu, Yi Qin, Wenyan Xu, Wensheng Liu, Xianjun Yu, and Jin Xu. 2018. "The Impact of Cancer-Associated Fibroblasts on Major Hallmarks of Pancreatic Cancer." *Theranostics* 8 (18): 5072–87. <https://doi.org/10.7150/thno.26546>.
- Tahira, Ana C, Márcia S Kubrusly, Michele F Faria, Bianca Dazzani, Rogério S Fonseca, Vinicius Maracaja-Coutinho, Sergio Verjovski-Almeida, Marcel CC Machado, and Eduardo M Reis. 2011a. "Long Noncoding Intronic RNAs Are Differentially Expressed in Primary and Metastatic Pancreatic Cancer."

- Molecular Cancer* 10 (1): 141. <https://doi.org/10.1186/1476-4598-10-141>.
- . 2011b. “Long Noncoding Intronic RNAs Are Differentially Expressed in Primary and Metastatic Pancreatic Cancer.” *Molecular Cancer* 10 (1): 141. <https://doi.org/10.1186/1476-4598-10-141>.
- Tang, Jianing, Deguang Kong, Qiuxia Cui, Kun Wang, Dan Zhang, Yan Gong, and Gaosong Wu. 2018. “Prognostic Genes of Breast Cancer Identified by Gene Co-Expression Network Analysis.” *Frontiers in Oncology* 8: 374. <https://doi.org/10.3389/fonc.2018.00374>.
- Theobald, Benjamin, Kathy Bonness, Alla Musiyenko, Joel F. Andrews, Gudrun Urban, Xizhong Huang, Nicholas M. Dean, and Richard E. Honkanen. 2013. “Suppression of Ser/Thr Phosphatase 4 (PP4C/ PPP4C) Mimics a Novel Post-Mitotic Action of Fostriecin, Producing Mitotic Slippage Followed by Tetraploid Cell Death.” *Molecular Cancer Research* 11 (8): 845–55. <https://doi.org/10.1158/1541-7786.mcr-13-0032>.
- Therneau, Terry M., and Patricia M. Grambsch. 2000. *Modeling Survival Data : Extending the Cox Model*. Springer. <https://cran.r-project.org/web/packages/survival/citation.html>.
- Therneau, Terry M. n.d. “Survival Analysis [R Package Survival Version 2.42-3].” Accessed June 21, 2018. <https://cran.r-project.org/web/packages/survival/index.html>.
- Torres, Carolina, and Paul J. Grippo. 2018. “Pancreatic Cancer Subtypes: A Roadmap for Precision Medicine.” *Annals of Medicine* 50 (4): 277–87. <https://doi.org/10.1080/07853890.2018.1453168>.
- Trapnell, Cole, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, and Lior Pachter. 2012. “Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks.” *Nature Protocols* 7 (3): 562–78. <https://doi.org/10.1038/nprot.2012.016>.
- Tsai, M.-C., O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal, and H. Y. Chang. 2010. “Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes.” *Science* 329 (5992): 689–93. <https://doi.org/10.1126/science.1192002>.
- Tye, Coralee E., Joseph R. Boyd, Natalie A. Page, Michelle M. Falcone, Janet L. Stein, Gary S. Stein, and Jane B. Lian. 2018. “Regulation of Osteogenesis

- by Long Noncoding RNAs: An Epigenetic Mechanism Contributing to Bone Formation.” *Connective Tissue Research* 59 (sup1): 35–41. <https://doi.org/10.1080/03008207.2017.1412432>.
- Vincent, Audrey, Joseph Herman, Rich Schulick, Ralph H Hruban, and Michael Goggins. 2011. “Pancreatic Cancer.” *The Lancet* 378 (9791): 607–20. [https://doi.org/10.1016/S0140-6736\(10\)62307-0](https://doi.org/10.1016/S0140-6736(10)62307-0).
- Vishnubalaji, Radhakrishnan, Hibah Shaath, Eyad Elkord, and Nehad M. Alajez. 2019. “Long Non-Coding RNA (LncRNA) Transcriptional Landscape in Breast Cancer Identifies LINC01614 as Non-Favorable Prognostic Biomarker Regulated by TGF β and Focal Adhesion Kinase (FAK) Signaling.” *Cell Death Discovery* 5 (1): 109. <https://doi.org/10.1038/s41420-019-0190-6>.
- Waddell, Nick Nicola Nick, Marina Pajic, Ann-Marie Patch, David K. Chang, Karin S. Kassahn, Peter Bailey, Amber L. Johns, et al. 2015. “Whole Genomes Redefine the Mutational Landscape of Pancreatic Cancer.” *Nature* 518 (7540): 495–501. <https://doi.org/10.1038/nature14169>.
- Wan, Ledong, Wenying Yu, Enhui Shen, Wenjie Sun, Yuan Liu, Jianlu Kong, Yihua Wu, et al. 2019. “SRSF6-Regulated Alternative Splicing That Promotes Tumour Progression Offers a Therapy Target for Colorectal Cancer.” *Gut* 68 (1): 118–29. <https://doi.org/10.1136/gutjnl-2017-314983>.
- Wang, Eric T., Rickard Sandberg, Shujun Luo, Irina Khrebtukova, Lu Zhang, Christine Mayr, Stephen F. Kingsmore, Gary P. Schroth, and Christopher B. Burge. 2008. “Alternative Isoform Regulation in Human Tissue Transcriptomes.” *Nature* 456 (7221): 470–76. <https://doi.org/10.1038/nature07509>.
- Wang, Feng, Zhongqiong Tang, Honglian Shao, Jun Guo, Tao Tan, Yang Dong, and Lianbing Lin. 2018. “Long Noncoding RNA HOTTIP Cooperates with CCCTC-Binding Factor to Coordinate HOXA Gene Expression.” *Biochemical and Biophysical Research Communications* 500 (4): 852–59. <https://doi.org/10.1016/j.bbrc.2018.04.173>.
- Wang, Weiwei, Chengguo Wei, Pan Li, Li Wang, Wencai Li, Kuisheng Chen, Jianying Zhang, Weijia Zhang, and Guozhong Jiang. 2018. “Integrative Analysis of mRNA and LncRNA Profiles Identified Pathogenetic LncRNAs in Esophageal Squamous Cell Carcinoma.” *Gene* 661 (June): 169–75. <https://doi.org/10.1016/J.GENE.2018.03.066>.

- Wang, Yan, Jing Liu, B O Huang, Yan-Mei Xu, Jing Li, Lin-Feng Huang, Jin Lin, et al. 2015. "Mechanism of Alternative Splicing and Its Regulation." *Biomedical Reports* 3 (2): 152–58. <https://doi.org/10.3892/br.2014.407>.
- Wang, Yun, Jia-Huan Lu, Ying-Nan Wang, and Huai-Qiang Ju. 2019. "IDDF2019-ABS-0268 Linc00920 Contributes to the Maintenance of mRNA Stabilizer IGF2BP2 in Colorectal Cancer." *Gut* 68 (Suppl 1): A26–27. <https://doi.org/10.1136/gutjnl-2019-iddfabstracts.52>.
- Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. "RNA-Seq: A Revolutionary Tool for Transcriptomics." *Nature Reviews Genetics* 10 (1): 57–63. <https://doi.org/10.1038/nrg2484>.
- Wartenberg, Martin, Silvia Cibi, Inti Zlobec, Erik Vassella, Serenella Eppenberger-Castori, Luigi Terracciano, Micha David Eichmann, et al. 2018. "Integrated Genomic and Immunophenotypic Classification of Pancreatic Cancer Reveals Three Distinct Subtypes with Prognostic/Predictive Significance." *Clinical Cancer Research* 24 (18): 4444–54. <https://doi.org/10.1158/1078-0432.CCR-17-3401>.
- Wei, Lusheng, Huilin Ye, Guolin Li, Yuanting Lu, Quanbo Zhou, Shangyou Zheng, Qing Lin, Yimin Liu, Zhihua Li, and Rufu Chen. 2018. "Cancer-Associated Fibroblasts Promote Progression and Gemcitabine Resistance via the SDF-1/SATB-1 Pathway in Pancreatic Cancer." *Cell Death & Disease* 9 (11): 1065. <https://doi.org/10.1038/s41419-018-1104-x>.
- Weinberg, Robert A. (Robert Allan). 2007. *The Biology of Cancer*. 2^a. https://books.google.com.br/books?hl=pt-BR&lr=&id=3WebDwAAQBAJ&oi=fnd&pg=PR18&dq=related:nPmlaUgjj00J:scholar.google.com/&ots=HUYxez5f8S&sig=PI0157JYcSxJ2JpH0qrUS_3Ue0s#v=onepage&q&f=false.
- Weinstein, John N, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. 2013. "The Cancer Genome Atlas Pan-Cancer Analysis Project." *Nature Publishing Group*. Vol. 45. <https://doi.org/10.1038/ng.2764>.
- Weiswald, Louis-Bastien, Dominique Bellet, and Virginie Dangles-Marie. 2015. "Spherical Cancer Models in Tumor Biology." *Neoplasia (New York, N.Y.)* 17 (1): 1–15. <https://doi.org/10.1016/j.neo.2014.12.004>.
- Wen, Xin, Xinran Tang, Yingqin Li, Xianyue Ren, Qingmei He, Xiaojing Yang,

- Jian Zhang, Yaqin Wang, Jun Ma, and Na Liu. 2016. "Microarray Expression Profiling of Long Non-Coding RNAs Involved in Nasopharyngeal Carcinoma Metastasis." *International Journal of Molecular Sciences* 17 (11): 1956. <https://doi.org/10.3390/ijms17111956>.
- Weng, Yuan-Chi, Jin Ma, Jun Zhang, and Jian-Cheng Wang. 2019. "Long Non-Coding RNA LINC01133 Silencing Exerts Antioncogenic Effect in Pancreatic Cancer through the Methylation of DKK1 Promoter and the Activation of Wnt Signaling Pathway." *Cancer Biology & Therapy* 20 (3): 368–80. <https://doi.org/10.1080/15384047.2018.1529110>.
- Wickramasinghe, Saumya, Angela Cánovas, Gonzalo Rincón, and Juan F. Medrano. 2014. "RNA-Sequencing: A Tool to Explore New Frontiers in Animal Genetics." *Livestock Science* 166 (August): 206–16. <https://doi.org/10.1016/J.LIVSCI.2014.06.015>.
- Witkiewicz, Agnieszka K., Elizabeth A. McMillan, Uthra Balaji, GuemHee Baek, Wan-Chi Lin, John Mansour, Mehri Mollae, et al. 2015. "Whole-Exome Sequencing of Pancreatic Cancer Defines Genetic Diversity and Therapeutic Targets." *Nature Communications* 6 (1): 6744. <https://doi.org/10.1038/ncomms7744>.
- Wood, Laura D., and Ralph H. Hruban. 2012. "Pathology and Molecular Genetics of Pancreatic Neoplasms." *The Cancer Journal* 18 (6): 492–501. <https://doi.org/10.1097/PPO.0b013e31827459b6>.
- WR Brugge, GY Lauwers, D Sahani, C Fernandez-del Castillo, AL Warshaw. 2004. "Cystic Neoplasms of the Pancreas." *N. Engl. J. Med.* 351: 1218–26.
- Xue, Zhihong, Scott Hennelly, Boryana Doyle, Arune A. Gulati, Irina V. Novikova, Karissa Y. Sanbonmatsu, and Laurie A. Boyer. 2016. "A G-Rich Motif in the LncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage." *Molecular Cell* 64 (1): 37–50. <https://doi.org/10.1016/J.MOLCEL.2016.08.010>.
- Yamamoto, Keisuke, Keisuke Tateishi, Yotaro Kudo, Mayumi Hoshikawa, Mariko Tanaka, Takuma Nakatsuka, Hiroaki Fujiwara, et al. 2016. "Stromal Remodeling by the BET Bromodomain Inhibitor JQ1 Suppresses the Progression of Human Pancreatic Cancer." *Oncotarget* 7 (38): 61469–84. <https://doi.org/10.18632/oncotarget.11129>.
- Yang, Fan, Huafeng Zhang, Yide Mei, and Mian Wu. 2014. "Reciprocal

- Regulation of HIF-1 α and LincRNA-P21 Modulates the Warburg Effect.” *Molecular Cell* 53 (1): 88–100. <https://doi.org/10.1016/j.molcel.2013.11.004>.
- Yang, Jue, and Hui Song. 2019. “Identification of Long Noncoding RNA RP11-169F17.1 and RP11-669N7.2 as Novel Prognostic Biomarkers of Stomach Adenocarcinoma Based on Integrated Bioinformatics Analysis.” *Epigenomics*, August, epi-2019-0115. <https://doi.org/10.2217/epi-2019-0115>.
- Yang, Shan-Zhong, Fei Xu, Tong Zhou, Xinyang Zhao, Jay M McDonald, and Yabing Chen. 2017. “The Long Non-Coding RNA HOTAIR Enhances Pancreatic Cancer Resistance to TNF-Related Apoptosis-Inducing Ligand.” *The Journal of Biological Chemistry* 292 (25): 10390–97. <https://doi.org/10.1074/jbc.M117.786830>.
- Yang, Xian-Zi, Tian-Tian Cheng, Qing-Jun He, Zi-Ying Lei, Jun Chi, Zhen Tang, Quan-Xing Liao, Hong Zhang, Li-Si Zeng, and Shu-Zhong Cui. 2018. “LINC01133 as CeRNA Inhibits Gastric Cancer Progression by Sponging MiR-106a-3p to Regulate APC Expression and the Wnt/ β -Catenin Pathway.” *Molecular Cancer* 17 (1): 126. <https://doi.org/10.1186/s12943-018-0874-1>.
- Yang, Xiaokun, Yadong Yang, Shuqian Tang, Hui Tang, Guihong Yang, Qiaoyu Xu, and Jinjin Wu. 2015. “EphB4 Inhibitor Overcome the Acquired Resistance to Cisplatin in Melanomas Xenograft Model.” *Journal of Pharmacological Sciences* 129 (1): 65–71. <https://doi.org/10.1016/J.JPHS.2015.08.009>.
- Yin, Qing-Fei, Ling-Ling Chen, and Li Yang. 2015. “Fractionation of Non-Polyadenylated and Ribosomal-Free RNAs from Mammalian Cells.” In , 69–80. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-1369-5_6.
- Yuan, Q, J Quackenbush, R Sultana, M Pertea, S L Salzberg, and C R Buell. 2001. “Rice Bioinformatics. Analysis of Rice Sequence Data and Leveraging the Data to Other Plant Species.” *Plant Physiology* 125 (3): 1166–74. <https://doi.org/10.1104/pp.125.3.1166>.
- Zang, Chongshuang, Feng-qi Nie, Qian Wang, Ming Sun, Wei Li, Jing He, Meiling Zhang, Kai-hua Lu, and Feng-qi Nie, Qian Wang, Ming Sun, Wei Li, Jing He, Meiling Zhang, Kai-hua Lu Chongshuang Zang. 2016. “Long Non-Coding RNA LINC01133 Represses KLF2, P21 and E-Cadherin Transcription

- through Binding with EZH2, LSD1 in Non Small Cell Lung Cancer.” *Oncotarget* 7 (10): 11696. <https://doi.org/10.18632/oncotarget.7077>.
- Zanoni, Michele, Filippo Piccinini, Chiara Arienti, Alice Zamagni, Spartaco Santi, Rolando Polico, Alessandro Bevilacqua, and Anna Tesei. 2016. “3D Tumor Spheroid Models for in Vitro Therapeutic Screening: A Systematic Approach to Enhance the Biological Relevance of Data Obtained.” *Scientific Reports* 6 (1): 19103. <https://doi.org/10.1038/srep19103>.
- Zhang, Bin, and Steve Horvath. 2005. “A General Framework for Weighted Gene Co-Expression Network Analysis.” *Statistical Applications in Genetics and Molecular Biology* 4 (1): Article17. <https://doi.org/10.2202/1544-6115.1128>.
- Zhang, J., J. Baran, A. Cros, J. M. Guberman, S. Haider, J. Hsu, Y. Liang, et al. 2011. “International Cancer Genome Consortium Data Portal--a One-Stop Shop for Cancer Genomics Data.” *Database* 2011 (0): bar026–bar026. <https://doi.org/10.1093/database/bar026>.
- Zhang, Xiao-Ou, Qing-Fei Yin, Ling-Ling Chen, and Li Yang. 2014. “Gene Expression Profiling of Non-Polyadenylated RNA-Seq across Species.” *Genomics Data* 2 (December): 237–41. <https://doi.org/10.1016/j.gdata.2014.07.005>.
- Zhang, Ying, Yan Miao, Muhe Shang, Mxin Liu, Ran Liu, Ench Pan, Yuepu Pu, and Lihong Yin. 2019. “LincRNA-P21 Leads to G1 Arrest by P53 Pathway in Esophageal Squamous Cell Carcinoma.” *Cancer Management and Research* Volume 11 (July): 6201–14. <https://doi.org/10.2147/CMAR.S197557>.
- Zhao, Linjie, Gaili Ji, Xiaobing Le, Chenlu Wang, Lian Xu, Min Feng, Yaguang Zhang, et al. 2017. “Long Noncoding RNA LINC00092 Acts in Cancer-Associated Fibroblasts to Drive Glycolysis and Progression of Ovarian Cancer.” *Cancer Research* 77 (6): 1369–82. <https://doi.org/10.1158/0008-5472.CAN-16-1615>.
- Zhao, Shanrong, Ying Zhang, Ramya Gamini, Baohong Zhang, and David von Schack. 2018. “Evaluation of Two Main RNA-Seq Approaches for Gene Quantification in Clinical RNA Sequencing: PolyA+ Selection versus RRNA Depletion.” *Scientific Reports* 8 (1): 4781. <https://doi.org/10.1038/s41598-018-23226-4>.
- Zhao, Wei, Xiaping He, Katherine A Hoadley, Joel S Parker, David Neil Hayes,

and Charles M Perou. 2014. “Comparison of RNA-Seq by Poly (A) Capture, Ribosomal RNA Depletion, and DNA Microarray for Expression Profiling.” *BMC Genomics* 15 (1): 419. <https://doi.org/10.1186/1471-2164-15-419>.

ZHOU, YU, BO GONG, ZHI-LIN JIANG, SHAN ZHONG, XING-CHAO LIU, KE DONG, HE-SHUI WU, HONG-JI YANG, and SHI-KAI ZHU. 2016. “Microarray Expression Profile Analysis of Long Non-Coding RNAs in Pancreatic Ductal Adenocarcinoma.” *International Journal of Oncology* 48 (2): 670–80. <https://doi.org/10.3892/ijo.2015.3292>.

8. Lista de anexos

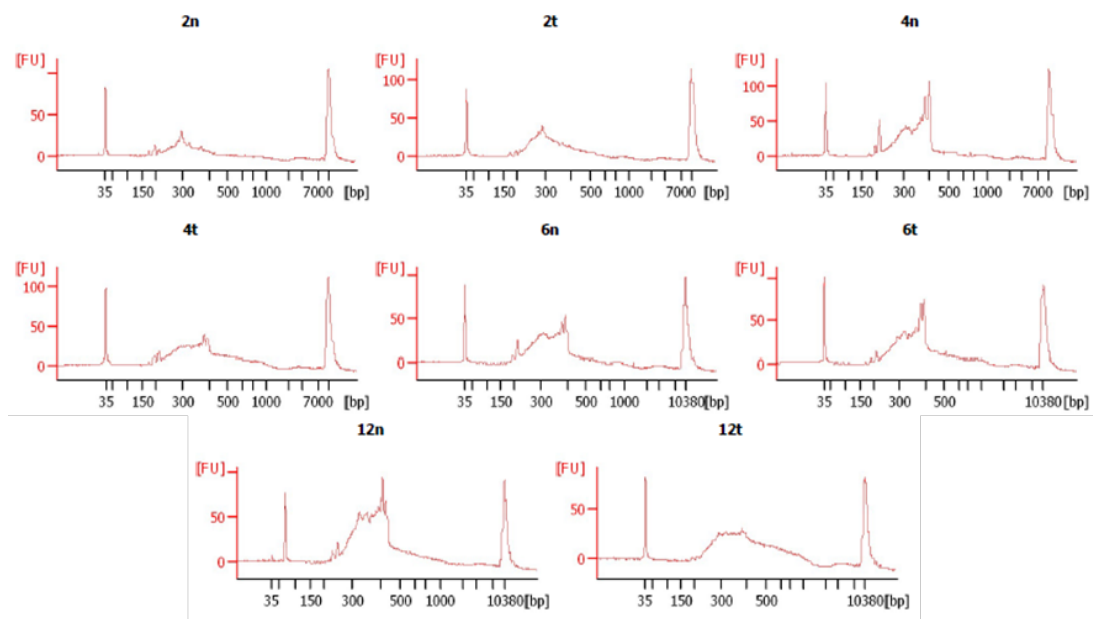
Anexo 1 – Perfil da eletroforese capilar das 28 bibliotecas de RNAseq.

Anexo 2 – Figuras e legendas do artigo em processo de submissão.

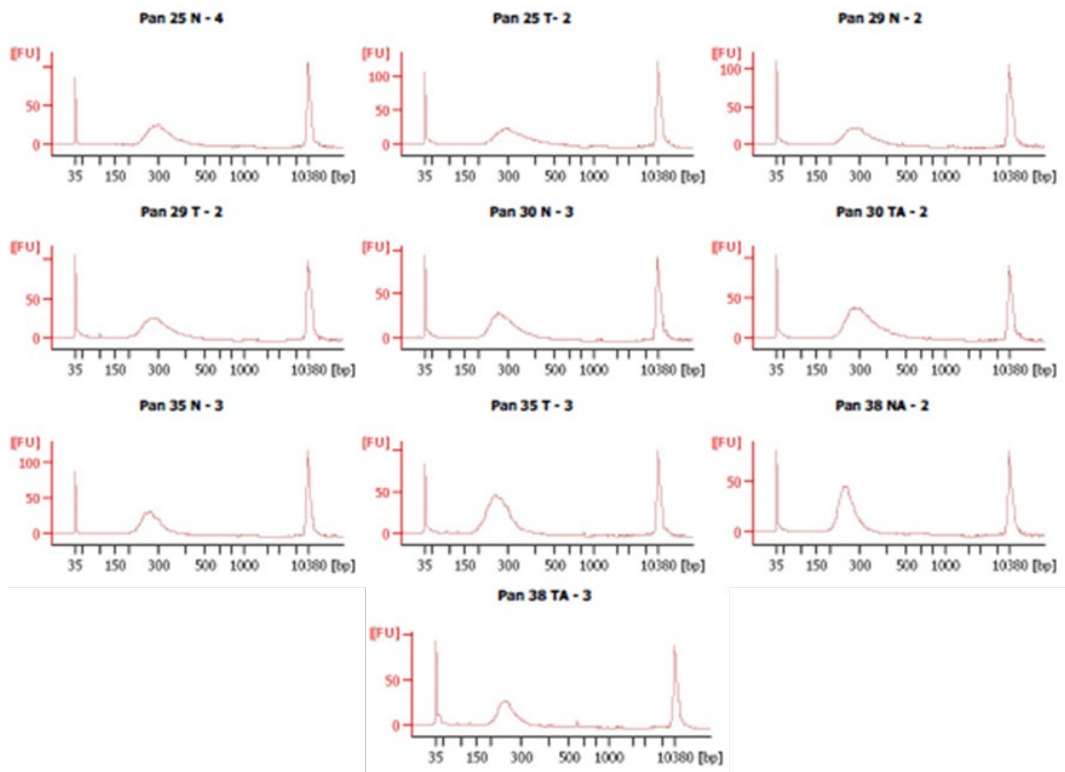
ANEXO 1 – Perfil de eletroforese capilar das 28 bibliotecas de RNAseq

As 28 bibliotecas de RNAseq (item 4.1) foram produzidas em três lotes utilizando-se o kit Illumina TruSeq Stranded Total RNA LT com Ribo-Zero Gold (Cat # RS-122-2001). Todas a partir do RNA extraído de amostras cirúrgicas de pâncreas, sendo 14 amostras pareadas (Tabela 1) com as amostras tumorais (T) e as respectivas amostras não tumorais pancreáticas adjacentes (N) foi realizada com sucesso, no tamanho médio adequado (300bp) para o sequenciamento em plataforma Illumina (Figura 6, item 4.1). As eletroforeses foram realizadas no aparelho Agilent 2100 Bioanalyzer do Centro Avançado de Tecnologias em Genômica (CATG) do IQUSP sob supervisão da especialista Dra. Layla Farage Martins.

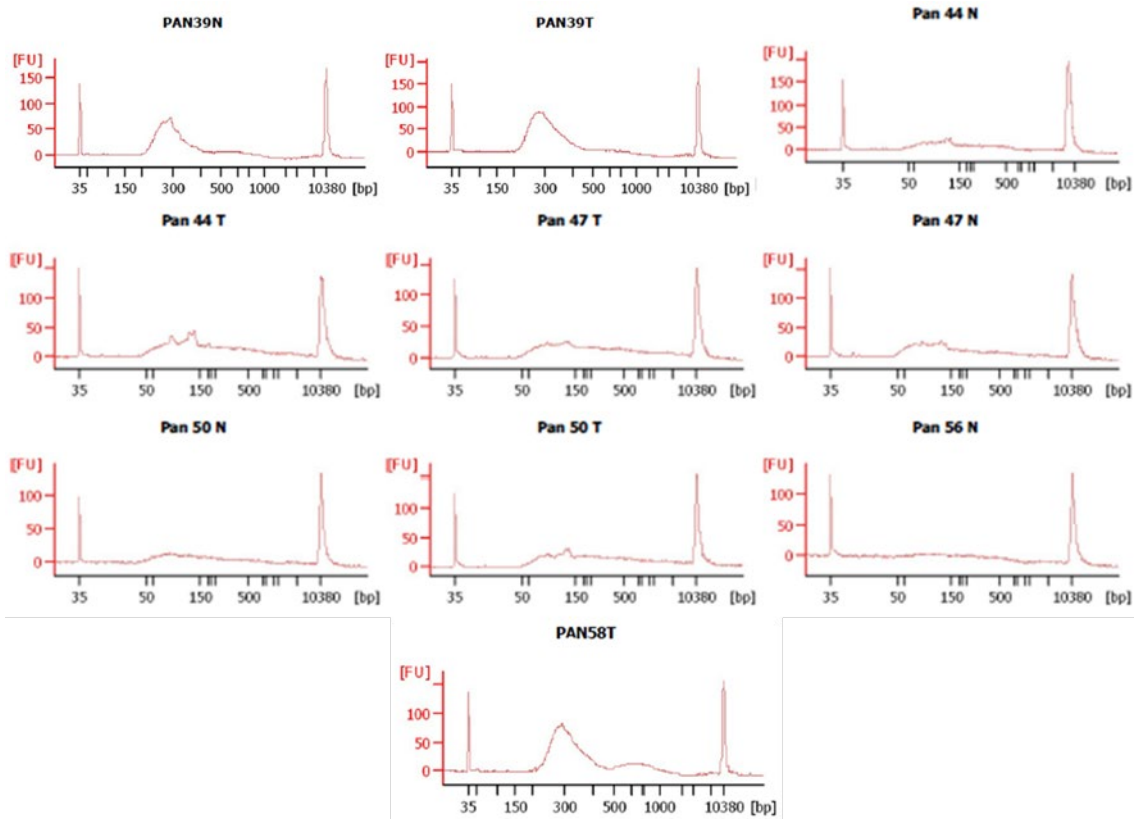
Perfil eletroforético capilar do lote 1 de bibliotecas processadas simultaneamente para RNAseq



Perfil eletroforético capilar do lote 2 de bibliotecas processadas simultaneamente para RNAseq



Perfil eletroforético capilar do lote 2 de bibliotecas processadas simultaneamente para RNAseq



Anexo 2 – Manuscrito em preparação para submissão no periódico BMC
Genome Medicine. (IF: 8.8 ISI)

**Transcriptional variation analysis of matched tumor/adjacent tissue
reveals pervasive alternative splicing and expression of long noncoding
RNAs in pancreatic adenocarcinoma**

Vinicius Ferreira da Paixão¹, Omar Julio Sosa², Diogo Pellegrina², Thalita
Bueno¹, Ester Bertoldi², Diogo Pessoa¹, Luís Bruno da Cruz e Alves-de-
Moraes¹, Victoria P. Oliveira¹, Ricardo A. C. Zevallos¹, Bianca Dazzani¹, Lilian
Cristina Russo Vieira¹, Fabio Forti¹, João Eduardo Ferreira³, José Jukemura⁴,
Marcel C C Machado⁴, Helano Freitas⁵, Maria Dirlei Begnami⁵, João C. Setubal¹,
Daniela S. Basseres¹, Eduardo M. Reis¹

¹Departamento de Bioquímica, Instituto de Química, Universidade de São
Paulo, São Paulo, SP, Brazil.

²Programa Interunidades de Pós-Graduação em Bioinformática, Universidade
de São Paulo, São Paulo, SP, Brazil

³Departamento de Ciência da Computação, Instituto de Matemática e
Estatística, Universidade de São Paulo, São Paulo, SP, Brazil.

⁴Departamento de Emergências Clínicas, Faculdade de Medicina da
Universidade de São Paulo, São Paulo, SP, Brazil

⁵CIPE - AC Camargo Cancer Center, São Paulo, SP, Brazil.

Corresponding author:

Eduardo M. Reis

Universidade de São Paulo, Instituto de Química.

Av. Prof. Lineu Prestes, 748, Cidade Universitária

05508-900 - São Paulo – SP, Brazil

Telephone: +55-11-3091-2173

E-mail: emreis@iq.usp.br

Abstract

Background: Global transcriptome analysis of pancreatic ductal adenocarcinoma (PDAC) has been useful to identify gene expression changes that sustain malignant phenotypes and to describe molecular subtypes of the disease. Yet, most studies examined only tumor tissues and focused on the transcriptional changes of well-annotated protein-coding genes, and the repertoire of long noncoding RNAs (lncRNAs) and alternative splicing variants expressed in pancreatic tissues of PDAC cases is still incomplete.

Methods: We generated strand-oriented total RNA-seq data of patient-matched tumor and adjacent nonmalignant pancreatic tissues from 14 cases of PDAC and implemented a computational pipeline to reconstruct the pancreatic transcriptome and annotate novel lncRNAs and splicing variants. The expression of a subset of novel transcripts was validated by RT-qPCR and the contribution of lncRNAs with aberrant expression in PDAC to malignant phenotypes was

examined following siRNA knockdown in tumor cell lines. Gene coexpression network and functional enrichment analyses were used to infer biological processes exerted by oncogenic lncRNAs.

Results: The assembled pancreatic transcriptome comprised 10570 transcripts GENCODE annotated lncRNAs as well as 6187 unannotated splicing variants of known genes, 339 intergenic RNAs and 184 antisense RNAs. Most unannotated intergenic/antisense RNAs display low or absent coding potential and their structural and regulatory genomic features are consistent with them being *bonafide* cellular transcripts. Hundreds of novel splicing variants and intergenic/antisense lncRNAs differentially expressed in tumor tissues were identified, most being detected in publicly available PDAC datasets (TCGA and ICGC) and including transcripts significantly associated to patient outcome. Aberrant expression in PDAC of a subset of novel and known lncRNAs was independently confirmed in patient samples, xenografts and cell lines. siRNA-mediated knockdown of lncRNAs LINC01559, LINC01133, CCAT1, LINC0920 and UCA1 reduced cell proliferation (excepted for LINC0920), migration and invasion in vitro. Gene coexpression network analysis associated lncRNAs to diverse biological processes in PDAC, such as cell adhesion, protein O-glycosylation and DNA repair. UCA1 is coexpressed with genes associated with DNA repair and UCA1 knockdown was shown to specifically deregulate the expression of genes in the same module and to negatively impact DNA repair in PDAC cells following damage induced by ionizing radiation.

Conclusions: Our study expands the catalog of splicing variants of known genes, intergenic and antisense long noncoding RNAs expressed in PDAC. A number of novel potential biomarkers for risk stratification of patients with pancreatic cancer

were identified. Based on a gene co-expression analysis we raise the possibility that lncRNAs deregulated in PDAC may participate in specific cellular processes relevant to sustain malignant phenotypes and demonstrate the participation of UCA1 in DNA repair.

Keywords: Pancreatic ductal adenocarcinoma, lncRNA, co-expression networks.

Figure 1. Unsupervised analysis of PDAC transcriptome data. Principal component (PCA, to the left) and hierarchical clustering (to the right) analysis of tumor (black) and non-tumor pancreatic tissue patient samples (white) was performed with the top 500 most variant transcripts among protein-coding (panel A), novel splicing isoforms (panel B), intergenic RNAs (panel C) and antisense RNAs (panel D).

Figure 2. Novel transcripts display structural and regulatory genomic features like known protein-coding and intergenic/antisense RNAs. Distributions of abundance (panel A), length (panel B), and number of exons (panel C) of the reconstructed transcripts. Distance distribution of GENCODE annotated promoter (panel D) and enhancer-like (panel E) regions relative to the TSS of known and novel transcripts. The distributions were significantly different (Kolmogorov-Smirnov test, $p < 0.001$) to the distribution observed for a randomly selected set of sequences. Transcript types are color labeled as follows: lilac, potentially novel isoform; red, novel intergenic RNA; blue: novel antisense RNA; black, GENCODE protein coding mRNA; orange: GENCODE lincRNA; light blue, GENCODE antisense RNA; gray: random sequences.

Figure 3. Detection in miTranscriptome and coding potential of PDAC novel transcripts. The overlap between unannotated transcripts detected in our study and the miTranscriptome catalog is shown in panel A (light green, transcripts detected only in this work; dark green, transcripts confirmed in miTranscriptome). The protein-coding potential of novel splicing variants of GENCODE transcripts and unannotated intergenic and antisense RNAs was inferred using the CPC or PhyloCSF programs (panel B) based on scores calculate with each approach

(see Methods for details). Panel C; diagram showing the intersection between non-coding/non-coding weak transcripts (coding potential score < 0) detected by CPC or PhyloCSF. Open Reading Frame (ORF) analysis was performed using TransDecoder (<https://transdecoder.github.io>), and ORFs with more than 100 residues were annotated. Similarity to the Swissprot database (based on a 1.0×10^{-5} significance threshold) was evaluated using BLASTP. Dark; transcripts without an ORF; Light red: ORFs without similarity to known proteins; gray: ORFs with significant similarity to Swissprot.

Figure 4. GENCODE annotated and novel transcripts differentially expressed in PDAC. Only transcripts with fold change $> |10|$ and $p\text{-adj} \leq 0.001$ were considered significant (see *Methods for details*). Heatmaps with the expression of a) protein coding ($n=669$); b) novel splicing isoforms ($n=171$); c) intergenic RNAs ($n=86$; 34 knowns, 52 novel) and d) antisense RNAs ($n=49$; 14 known, 35 novel) are shown. White bars: non-tumor pancreatic tissue patient samples; black bars: tumor samples. Z-score normalized expression values are shown.

Figure 5. Validation and prognostic potential of novel intergenic transcripts detected in PDAC. Panel A; genome structure of novel intergenic lncRNAs selected for RT-PCR validation. The arrows represent the exon border regions flanked by the PCR primers. Panel B; expression of selected candidates was tested by end-point RT-PCR in PDAC-derived cell lines (AsPC-1, MiaPaCa-2, BXPC-3, PANC-1, PDX-08) and fibroblasts (MRC-5) along with controls (gDNA, nontemplate control, no RT,). A 2% agarose gel electrophoresis is shown. Molecular size markers are shown in lane 1. The expected PCR amplicons were variably observed in different cell lines: TCONS00084995 (149bp); TCONS00009076 (145bp); TCONS00087289 (188bp); TCONS00006750 (114bp); TCONS00059572 (159bp); TCONS000036574 (153bp). Panel C; aberrant expression in PDAC of novel intergenic lncRNAs was confirmed by RT-qPCR in clinical samples of PDAC, nontumor adjacent tissue (NT) and patient-derived tumor xenografts (PDXs). Results are shown as \log_2 fold-change values relative to the expression in the NT samples. (*, $P \leq 0.05$) (**, $P \leq 0.01$) (***, $P \leq 0.001$). Panel D; heatmap with the expression of the intergenic lncRNA candidates in different PDAC cell lines as detected by RT-qPCR. Results are

expressed in fold-change values relative to the expression in PANC-1 cells. Panel E; Expression of novel intergenic transcripts was significantly (TCONS00036574, HR: 0.55 $p = 0.044$) or marginally (TCONS00085964, HR: 1.66, $p = 0.085$) associated to patient outcome in the ICGC dataset.

Figure 6. Expression and prognostic potential of known candidate lncRNAs in patient samples, patient-derived xenografts (PDXs) and PDAC cell lines. Panel A; expression of GENCODE annotated lncRNAs in clinical samples of PDAC and PDXs relative to adjacent non-tumor tissues (NT) was confirmed by RT-qPCR. For each transcript, the relative expression refers to the average expression in NT samples (which was set to 1) following normalization to a reference gene (see *Methods* for details). Panel B. relative expression of candidate lncRNAs in different PDAC cell lines measured by RT-qPCR (left panel; expressed as fold-change relative to the abundance of LINC02577 in AsPC-1) or RNA-seq (middle panel; FPKM values, from ArrayExpress ID: E-MTAB-2706). For each lncRNA, the Spearman correlation between the expression measured by RT-qPCR or RNA-seq is shown (right panel). Panel C; patients with higher levels of LINC01559, LINC02577 and UCA1 showed increased risk of cancer-associated death (HRs range 2.0-1.9, $p < 0.05$).

Figure 7. Impact of lncRNA knockdown in proliferation, migration and invasion of PDAC *in vitro*. AsPC-1 cells were transfected with siRNAs targeting LINC01559, LINC01133, CCAT1, LINC0920 or UCA1 or control siRNAs (siCTRL). Panel A; RT-qPCR analysis with the relative expression (% of depletion) of target lncRNAs in cells transfected with specific siRNAs compared to controls. Panel B; cell growth was monitored by real time imaging for 8 days after transfection with test (3 replicates, purple points) or control siRNAs (2 replicates, green points). Dotted vertical lines indicate the change of culture medium at day 2. Values collected at every 2,5 hrs were used to fit a linear function and statistically significant differences in the slope of the linear fits ($p < 0.001$) are indicated. Panel C; collected images were used to estimate cell motility by measuring the migration distance of individual cells (see *Methods* for details). For each condition, the migration distance of randomly selected cells was measured at every 20 min (left panel). The average migration of fifteen cells in test (3 replicates) and control (2

replicates) cells was calculated (right panel). Panel D; invasion assay. 1×10^5 transfected cells were placed in matrigel-coated culture inserts. After 24h, cells were stained and migration through the matrigel was evaluated by measuring the staining area using the ImageJ software. Data show mean \pm SD of three independent experiments. Statistically significant differences were inferred using Mann-Whitney test (proliferation and invasion) or unpaired Student's t-test (migration) and the following cut-offs: *, $p \leq 0.05$; **, $p \leq 0.01$; ****, $p \leq 10^{-4}$.

Figure 8. lncRNA UCA1 modulates the expression of DNA repair genes and siRNA-mediated UCA1 knockdown reduces DNA repair *in vitro*. Panel A; a pancreatic gene expression network module enriched in DNA repair protein coding genes and coexpressed lncRNAs is shown. Tumor upregulated (red bars) and downregulated (blue bars) transcripts are indicated (statistically significant differences are shown in dark colors). Pairwise connectivity values are depicted in gray scale. The upper green bar indicates the relative connectivity ("hubness") of the lncRNAs in the module. Panel B; relative expression of genes belonging to the DNA repair coexpression module (SPIDR and PPP4C) or from a distinct network module (MUC13) in AsPC-1 cells transfected with siRNA complementary to UCA1 (siUCA1) or control siRNAs (siCTRL). Significant changes in Wilcoxon rank test are indicated (*, $p \leq 0.05$). Panel C; cells transfected with UCA1 siRNAs or control siRNAs were gamma-irradiated and the relative levels of DNA damage were monitored before ("C") and at different time points (15, 30, 60, 120 min) by alkaline comet assay, expressed as OTM units ("olive-tail moment"). Data from 50 nuclei measured in three independent replicate experiments are shown. Significant differences were detected using ANOVA (*, $p \leq 0.05$) (****, $p \leq 10^{-4}$).

Figure 1

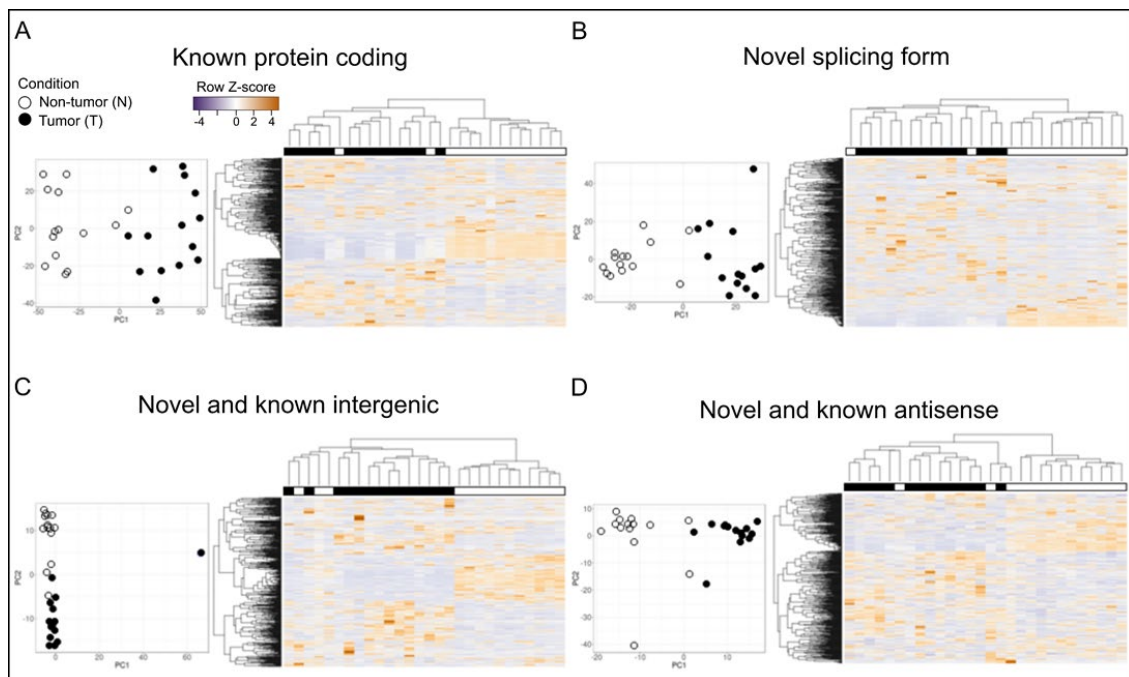


Figure 2

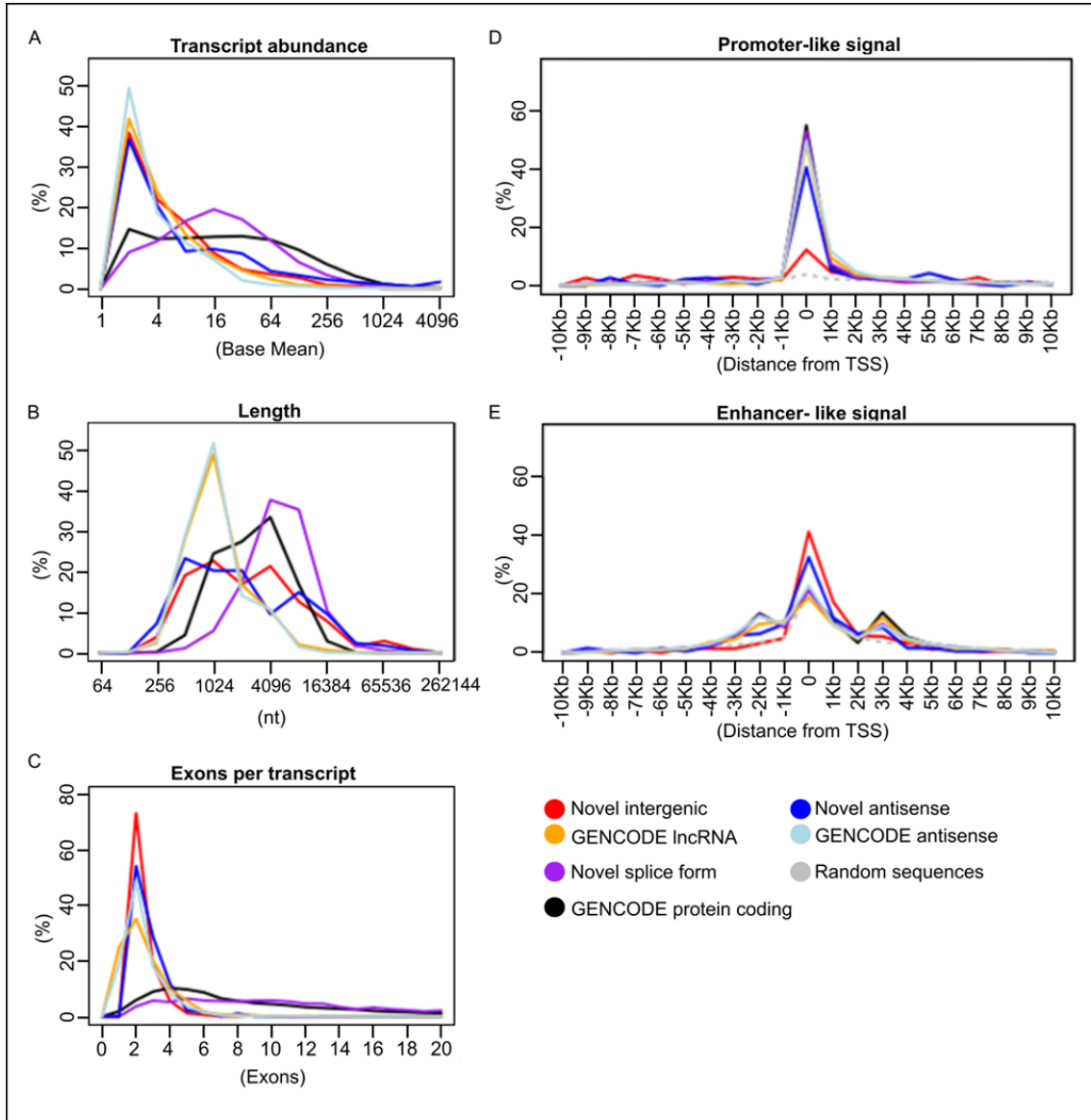


Figure 3

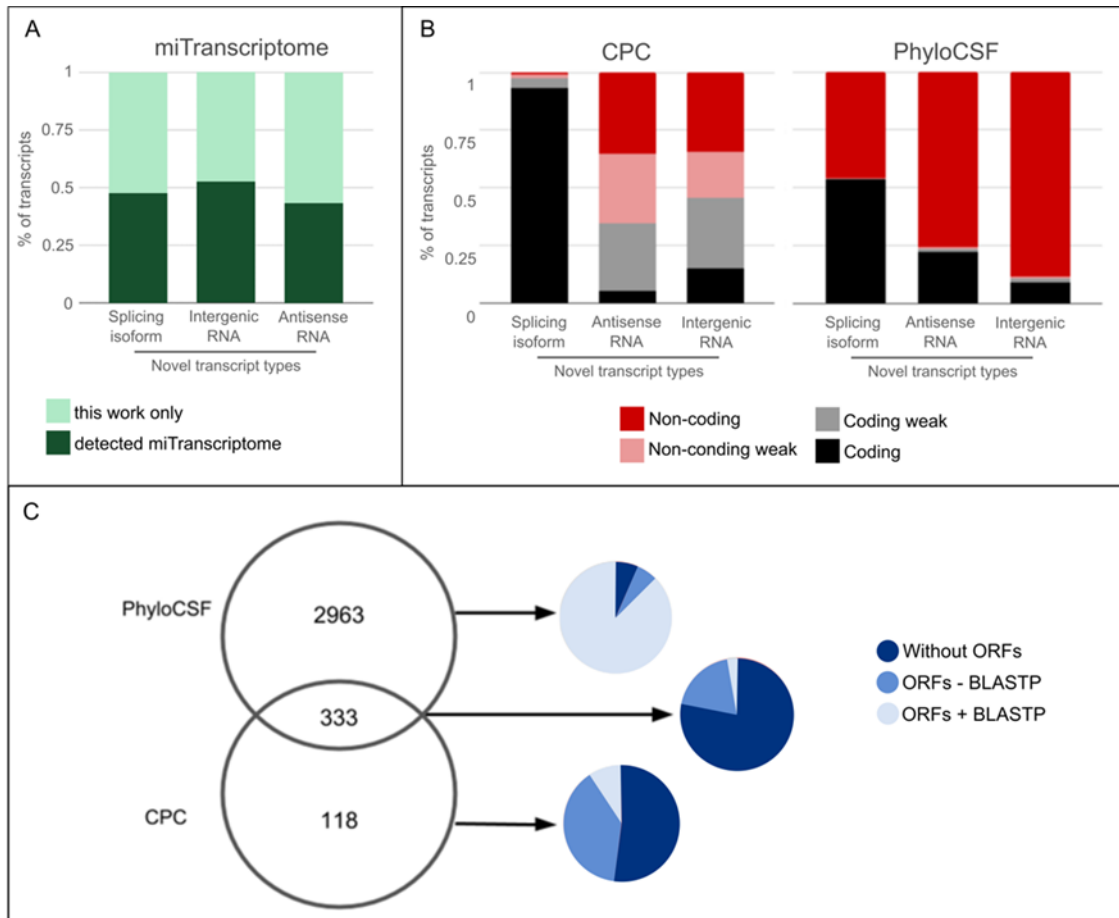


Figure 4.

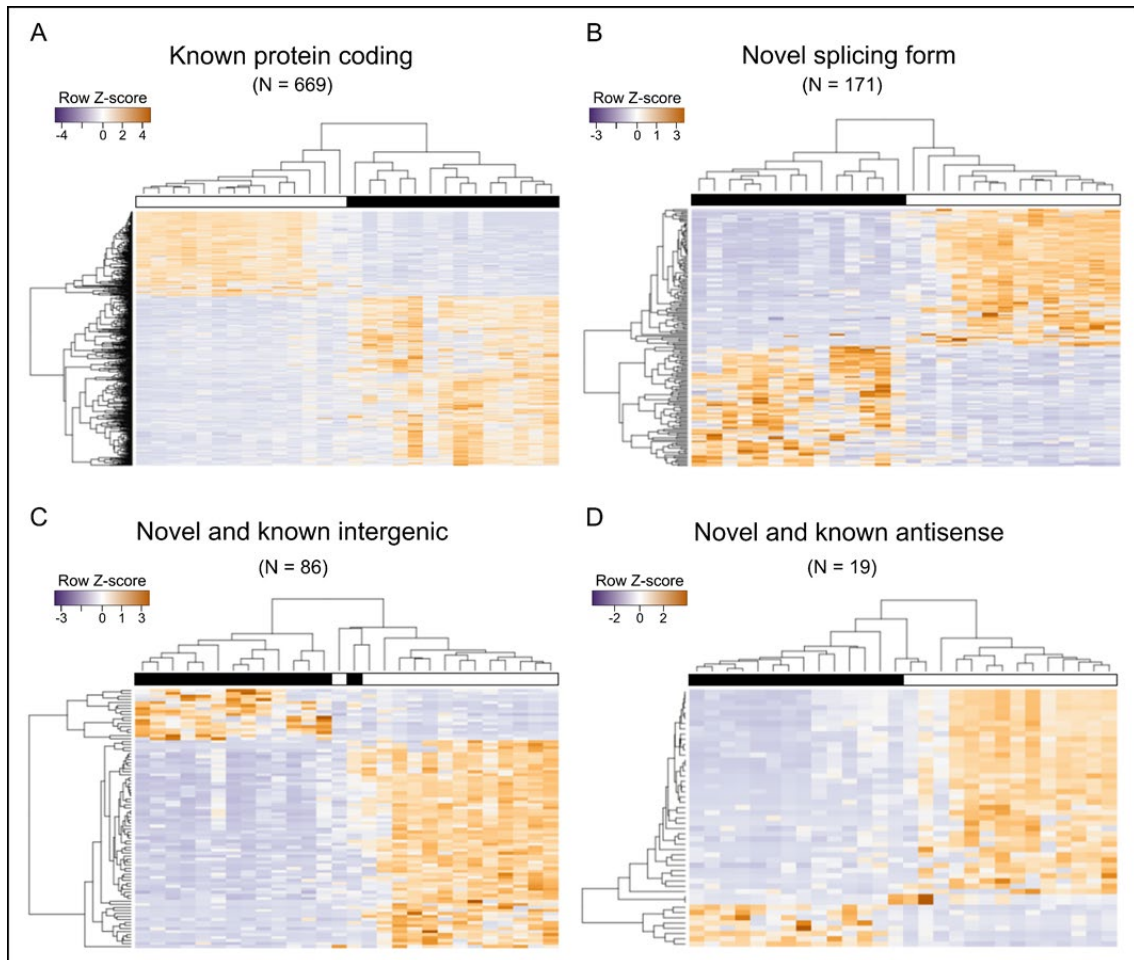


Figure 5

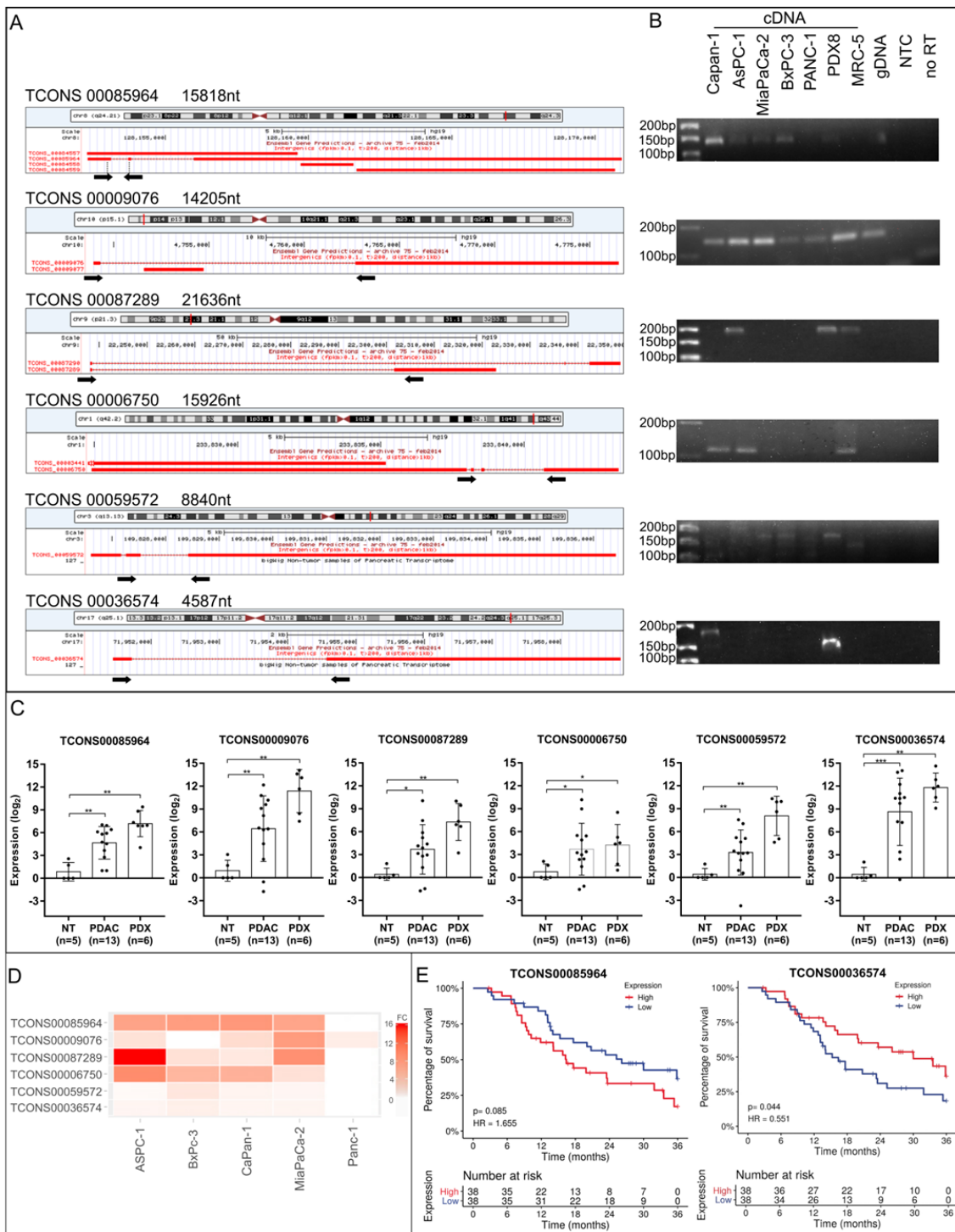


Figure 6

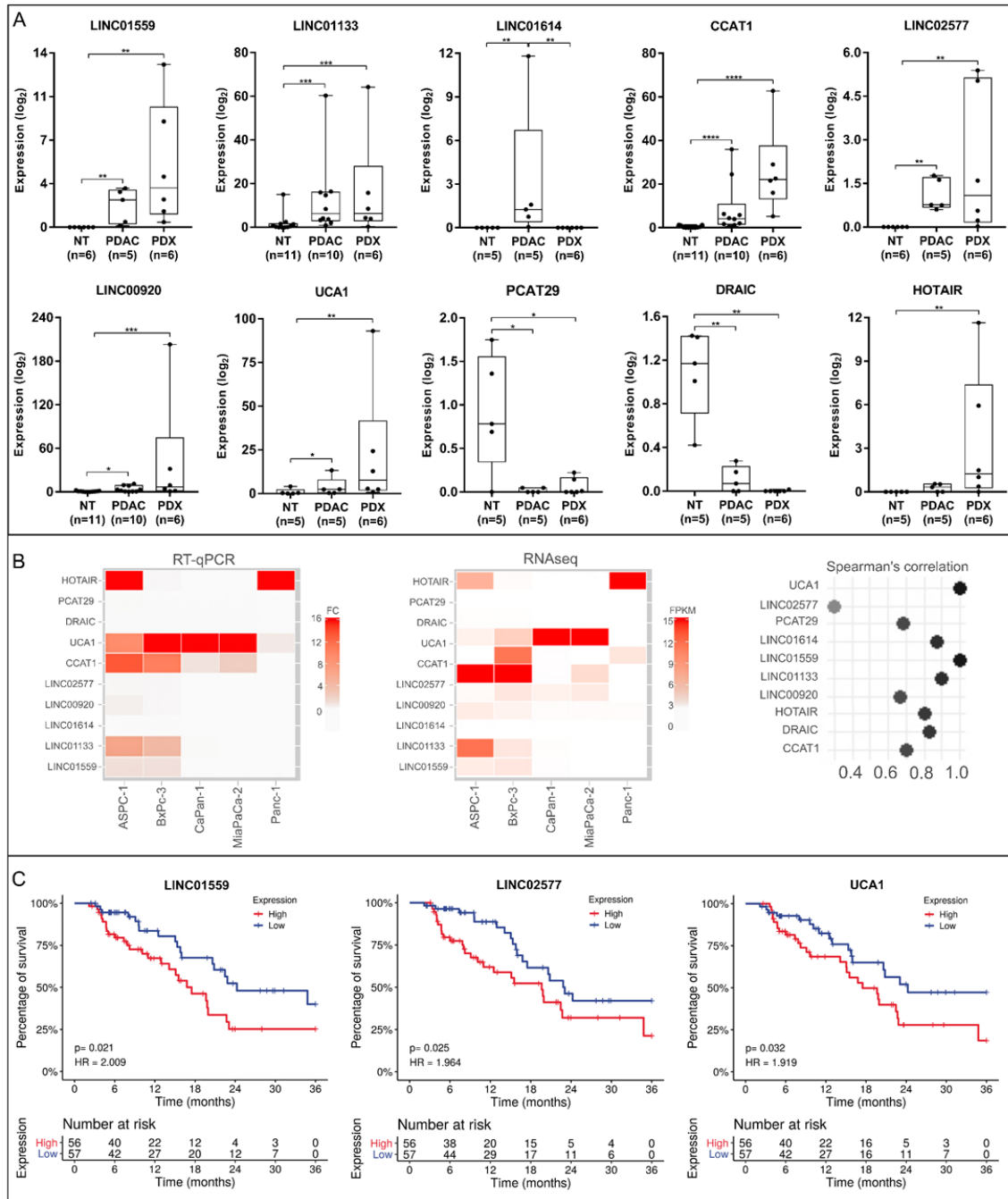


Figure 7

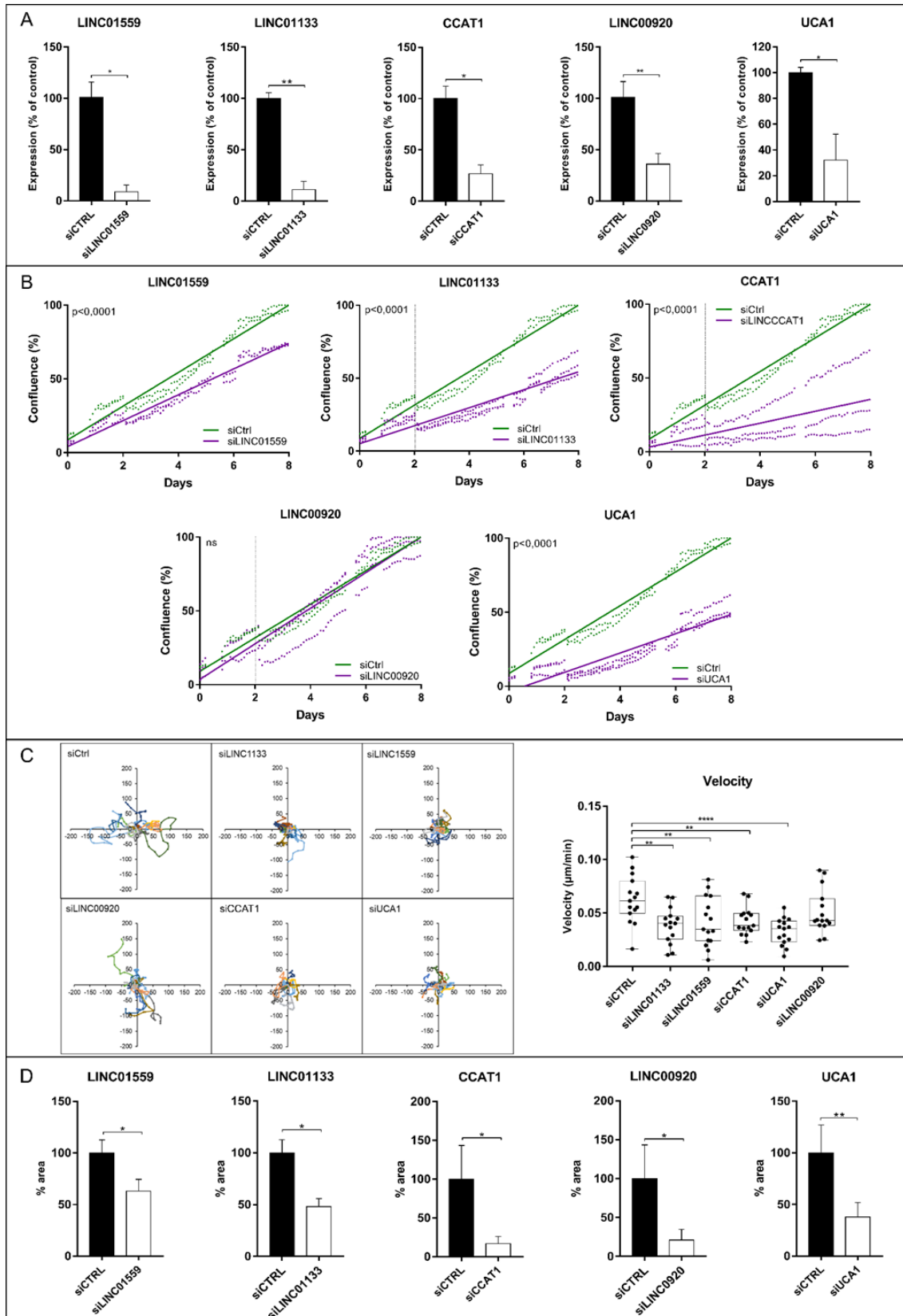
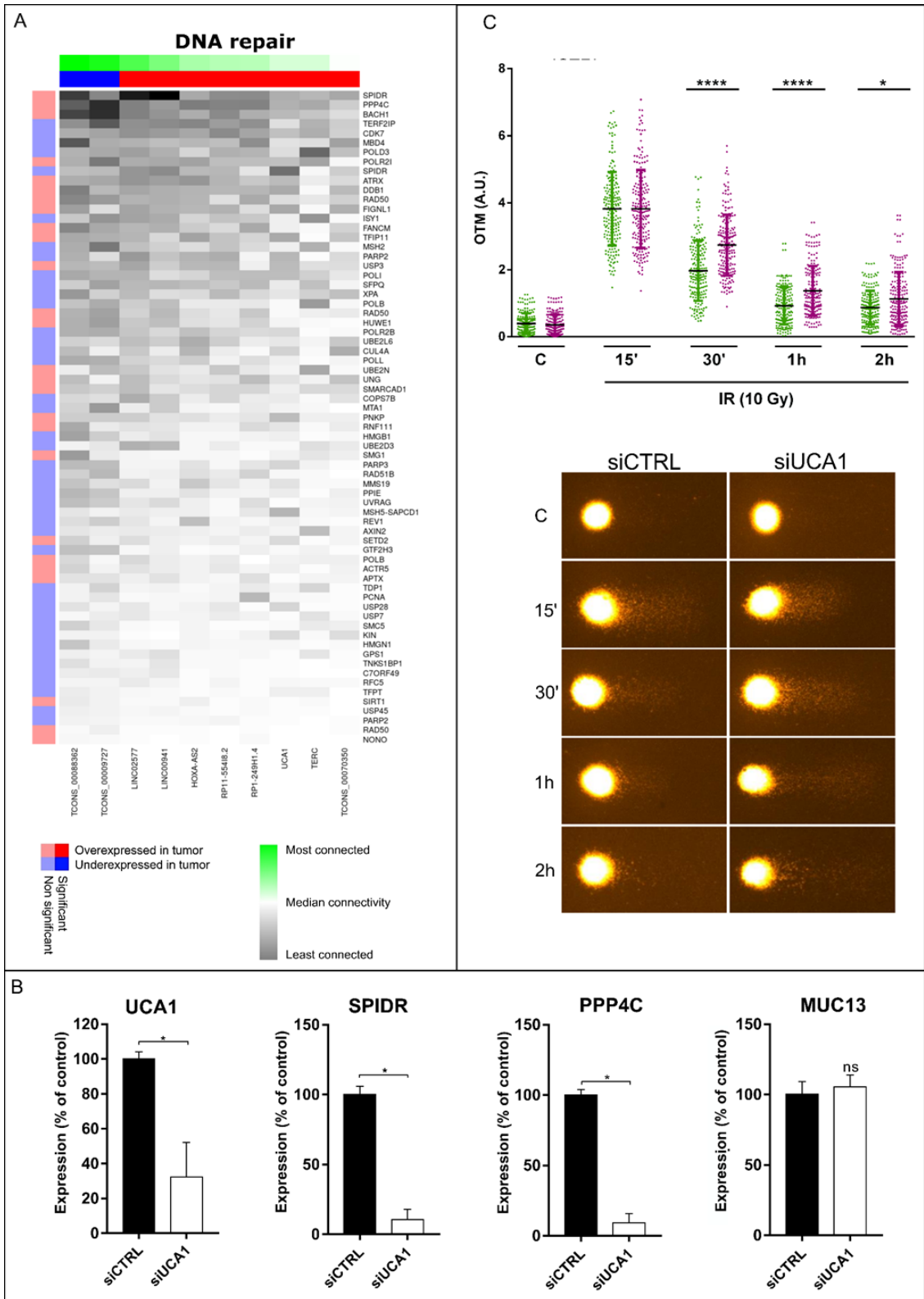


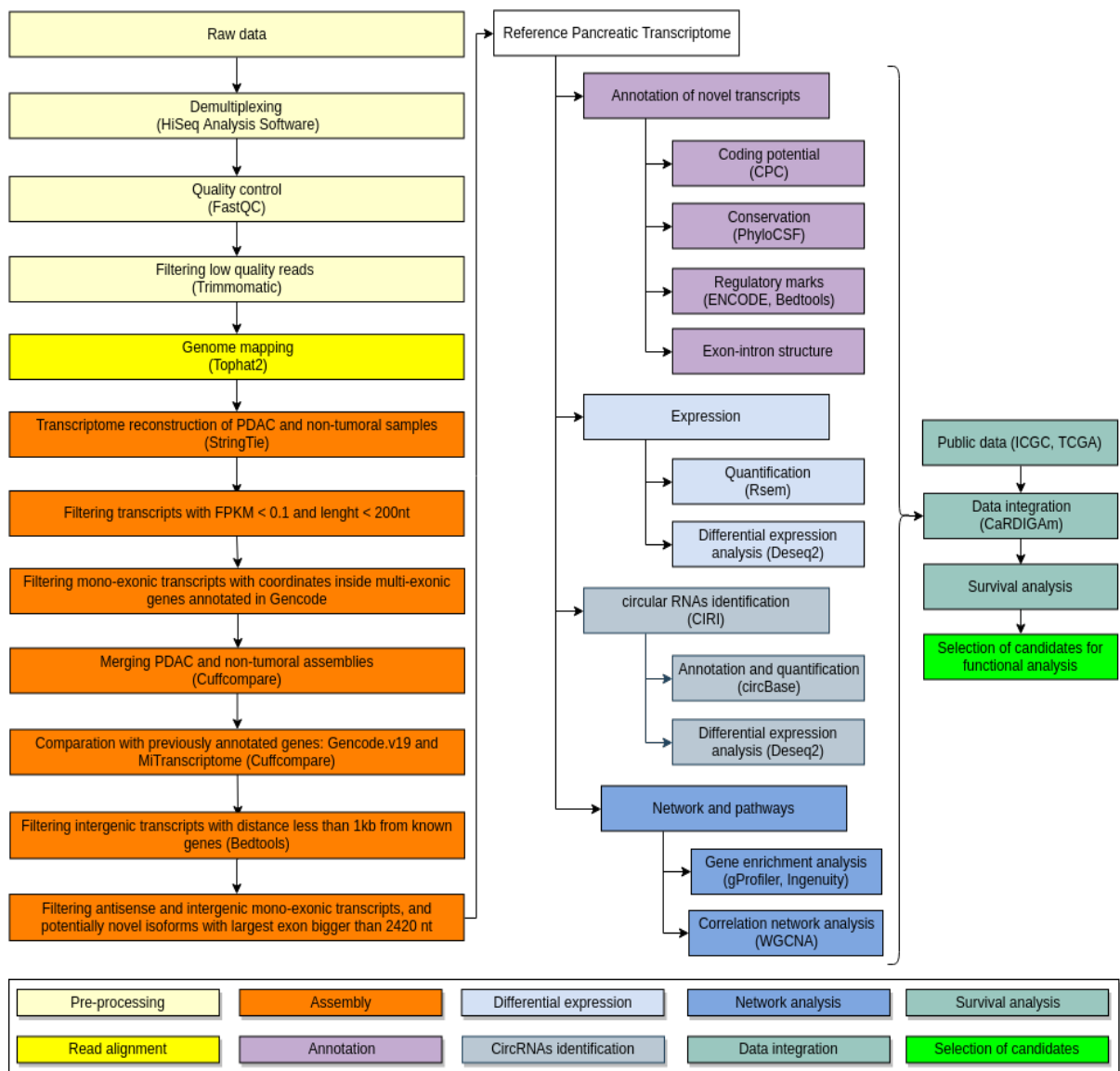
Figure 8



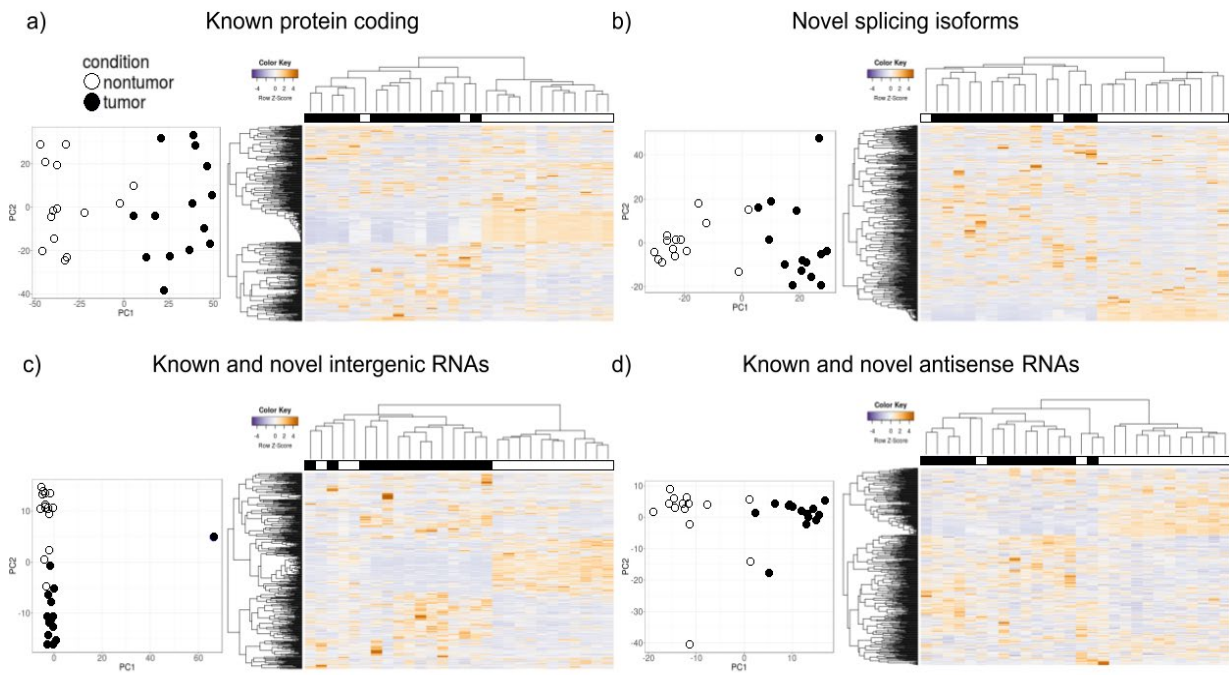
Legend to figures

Supplementary material

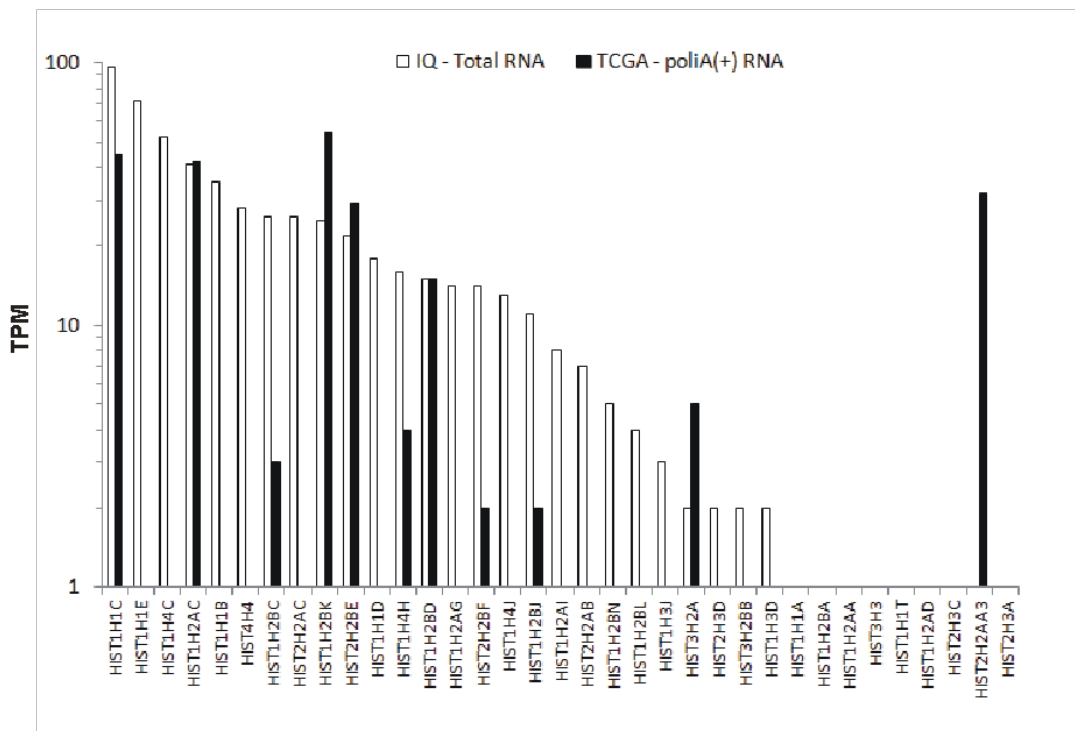
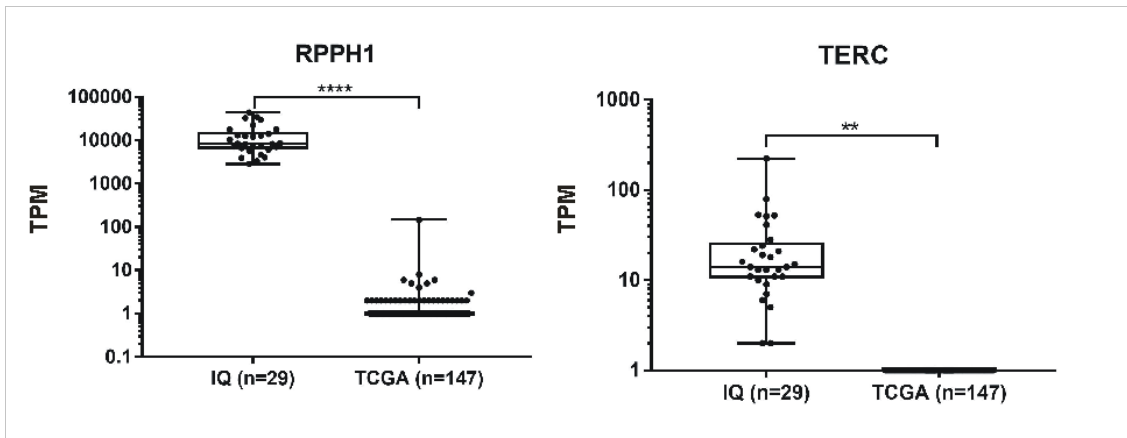
Supplementary Figure 1. The bioinformatics pipeline can be divided in nine modules: read pre-processing (light yellow), genome mapping (yellow), assembly (orange), annotation (violet), gene/transcript differential expression (light blue), network analysis (blue), circular RNA identification (gray), data integration (light green) and selection of candidates (green). Custom scripts written in Bash, R and Python were used to process input and output files from the standalone programs used in the pipeline.



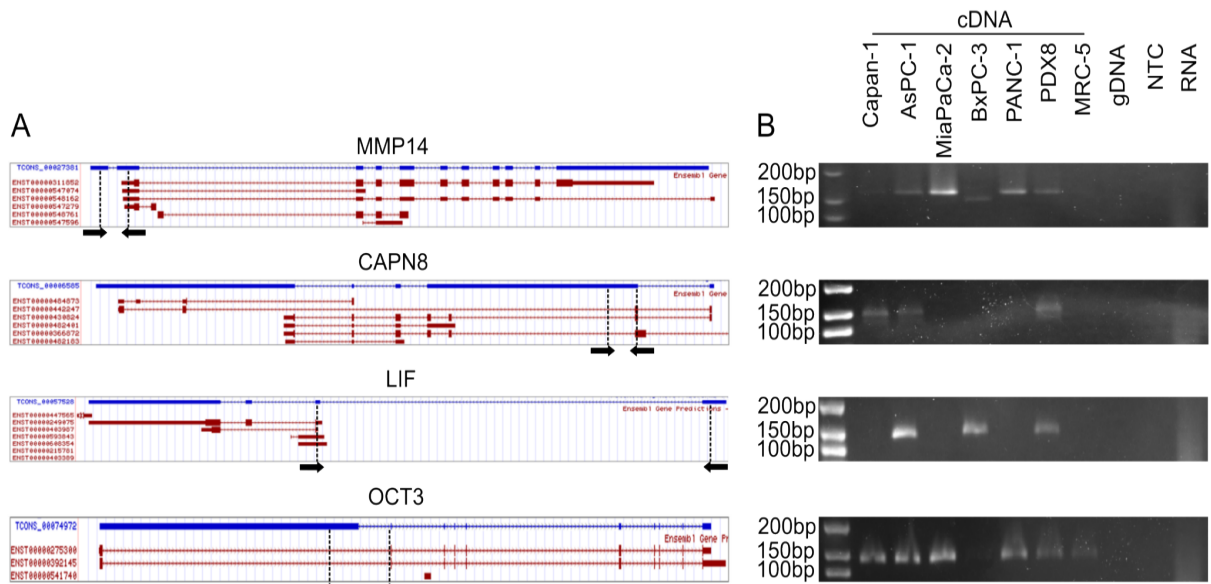
Supplementary Figure 2. Unsupervised analysis of PDAC transcriptome data. Principal component (PCA, to the left) and hierarchical clustering (to the right) analysis of tumour (black) and non-tumour pancreatic tissue patient samples (white) was performed with the top 500 most variant transcripts among protein-coding (panel A), novel splicing isoforms (panel B), intergenic RNAs (panel C) and antisense RNAs (panel D).



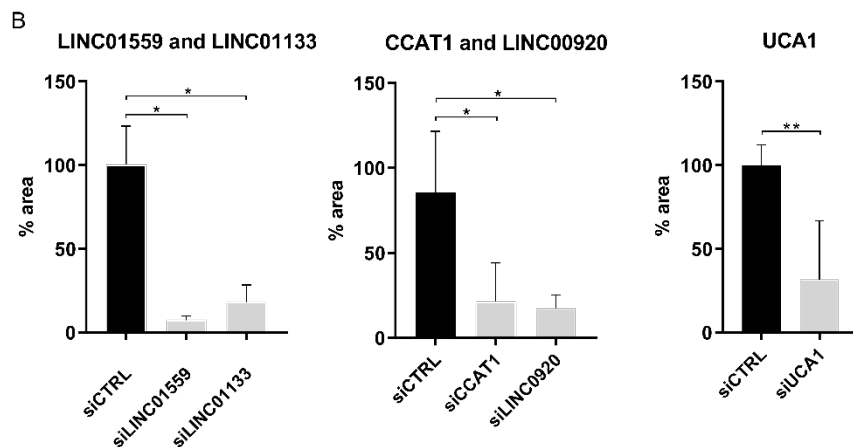
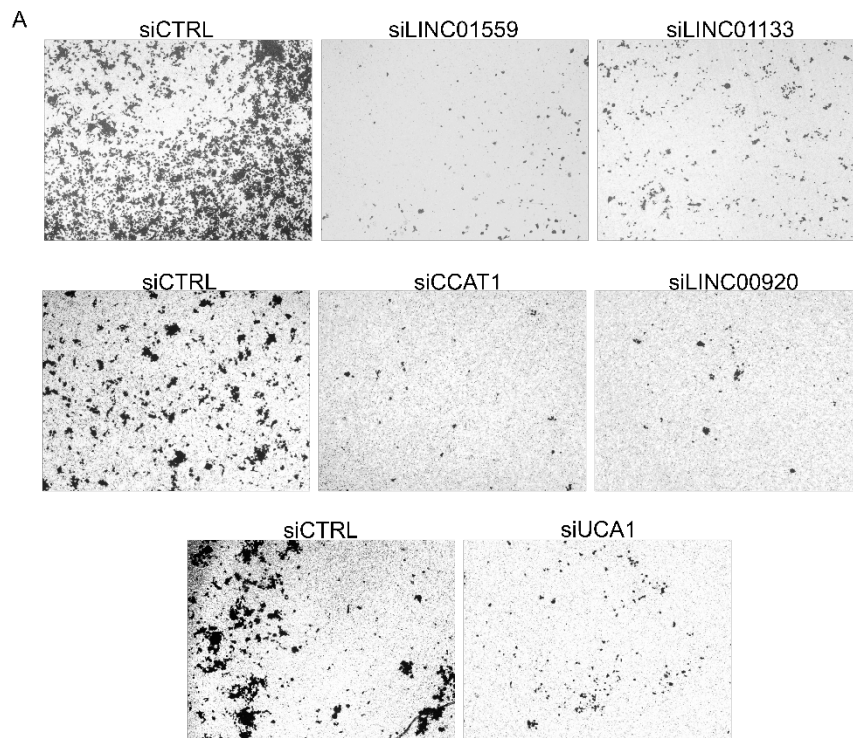
Supplementary Figure 3. rRNA depleted-RNAseq libraries from PDAC samples are enriched in nonpolyadenylated RNAs. Ribonuclease P RNA component H1 (RPPH1), telomerase RNA component (TERC) and several histone genes are detected at high expression levels in the RNAseq libraries generated from rRNA depleted RNA generated in this study (n= 29, IQ) compared to poliA-selected RNA (n= 147 samples, TCGA)



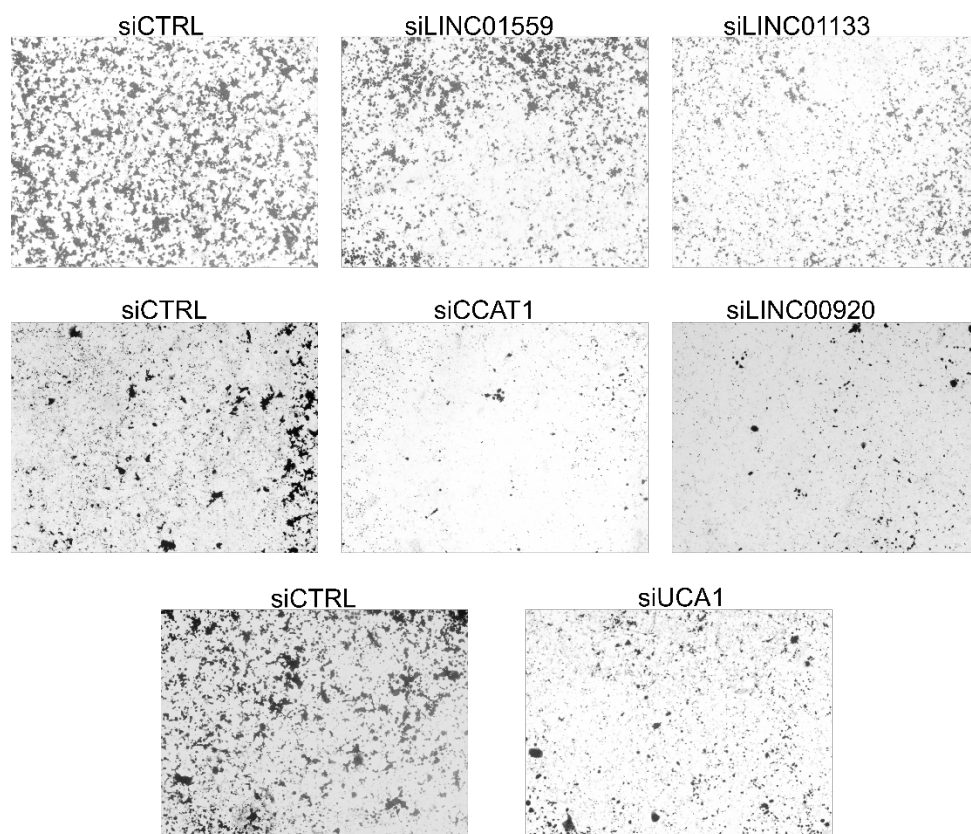
Supplementary Figure 4. AS validation. (A) Tracks generated in the Genome Browser to view of our assembly novel MMP14, CAPN8 and LIF AS transcripts. Tracks generated in the Genome Browser to view the new transcripts from our assembly. The arrows represent the exon border regions flanked by primers for PCR validation. (B) 2% agarose gel electrophoresis. At lane 1: Gene Ruler; lane 2: template, Capan-1 cDNA; lane 3: template, AsPC-1 cDNA; lane 4: template, MiaPaCa-2 cDNA; lane 5: template, Panc-1 cDNA; lane 6: template, BxPC-3 cDNA; lane 7: template, PDX cDNA; lane 8: template, MRC-5; lane 9: template, gDNA from CaPan-1; lane 10: NTC; lane 11: RNA from Capan-1. The expected amplicon PCR are observed



Supplementary Figure 5. Effect of lncRNA knockdown in Transwell migration assay. AsPC-1 cells (1x10⁵) transfected with siRNAs complementary to lncRNAs LINC01559, LINC01133, CCAT1, LINC00920, UCA1 or control siRNAs (siCTRL) were transferred to transwell plates in chemotatic conditions and counted after 24 hrs (see Methods for details). (A) Representative images of results from 3 replicate assays are shown. (B) (* *P*-value ≤ 0,05; ** *P*-value ≤ 0,01).

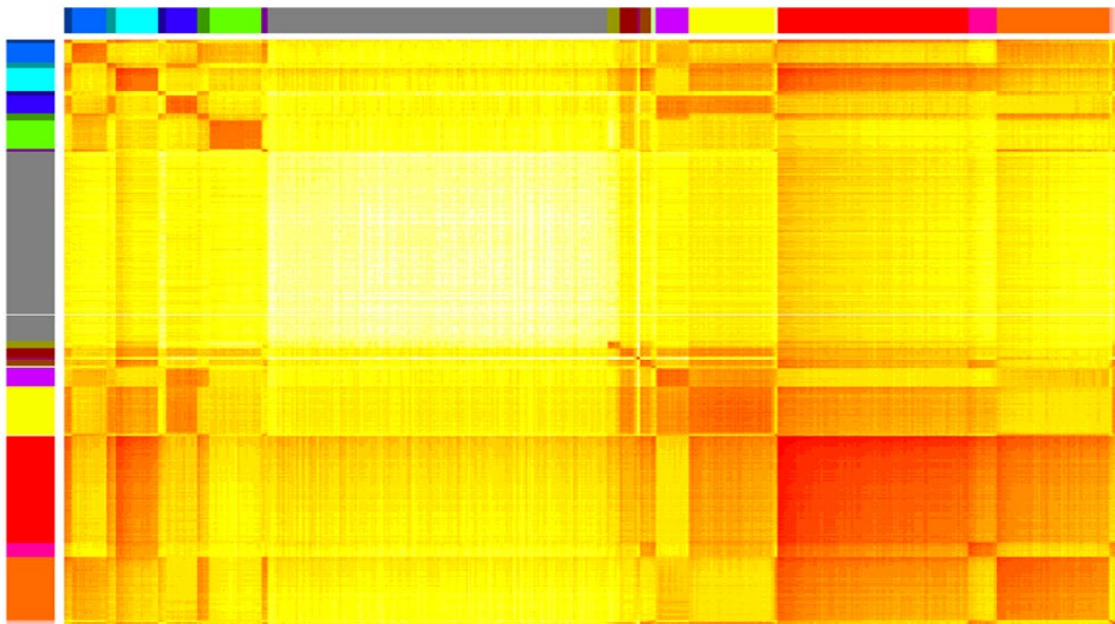


Supplementary Figure 6. Invasion assay. AsPC-1 cells (1×10^5) transfected with siRNAs complementary to lncRNAs LINC01559, LINC01133, CCAT1, LINC00920, UCA1 or control siRNAs (siCTRL) were transferred to transwell plates in chemotactic conditions and counted after 24 hrs (see Methods for details). Representative images of results from 3 replicate assays are shown



Supplementary Figure 7. Protein-coding/lncRNA coexpression network in PDAC and nontumor adjacent tissue A; heatmap with the correlation values in the co-expression network assembled with WGCNA (see *Methods* for details). Twenty-three network modules were identified (coloured in the superior and lateral bars) and clustered by their co-expression pattern. B; Transcripts in each module (lower axis) were used in functional enrichment analysis and selected significantly enriched biological processes ($p < 0.05$) are indicated.

A



B

Absolute correlation with tumor expression

