UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

Programa de Pós Graduação em Ciências Biológicas (Bioquímica)

ANA CAROLINA TAHIRA

Análise da expressão de RNAs não codificadores longos em adenocarcinoma de pâncreas.

Versão original da tese defendida

SÃO PAULO

Data do Depósito na SPG:

14/02/2013

ANA CAROLINA TAHIRA

Análise da expressão de RNAs não codificadores longos em adenocarcinoma de pâncreas

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

Área de Concentração: Bioquímica
Orientador: Dr. Eduardo Moraes Rego Reis

São Paulo
2013

Ficha Catalográfica
Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Tahira, Ana Carolina
T128a Análise da expressão de RNAs não codificadores longos em adenocarcinoma de pâncreas / Ana Carolina Tahira. -- São Paulo, 2013.
207p.

Tese (doutorado) - Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica.
Orientador: Reis, Eduardo Moraes Rego

1. Expressão gênica 2. Genoma : Biologia molecular I. T. II. Reis, Eduardo Moraes Rego, orientador.

574.88 CDD

Ana Carolina Tahira

Análise da expressão de RNAs não codificadores longos em adenocarcinoma de pâncreas.

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

Aprovado em: _____

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

DEDICATÓRIA

Ao meu pai pelo apoio incondicional às minhas escolhas, apesar de não as compreender.

À minha mãe por sempre fazer parte da minha vida e de quem eu sou, apesar de não estar mais aqui.

Aos meus irmãos pela chatice.

À minha família por sempre estarem comigo, mesmo os que ainda não chegaram (Theo).

À minha irmã querida do coração.

A todos os meus amigos, que sempre estiveram comigo ao longo de toda essa jornada.

AGRADECIMENTOS

Primeiramente eu tenho que agradecer a oportunidade de realizar este trabalho e pelas experiências que passei durante este período, as quais farão parte da minha vida para sempre.

Agradeço especialmente às minhas alunas de iniciação científica (Michelle e Bianca) que me ajudaram a amadurecer durante o doutorado, além do desenvolvimento do trabalho. Muito obrigada, muito do que consegui foi graças a vocês.

Ao Prof. Dr. Eduardo Moraes Reis pela oportunidade de desenvolver essa tese de doutorado e pela orientação durante todos esses anos. Apesar das grandes discussões sempre conseguimos trabalhar bem juntos.

Ao Prof. Sergio Verjovski-Almeida, por ter sido o meu orientador na iniciação científica, mas também pelo compartilhamento da estrutura do laboratório. Agradeço-o também pelas suas críticas ao meu trabalho, as quais, apesar de muitas vezes serem duras, contribuíram para o desenvolvimento o meu desenvolvimento científico e da tese.

Aos meus amigos do laboratório um agradecimento muito especial por enfrentarem muitas marés ruins comigo:

À Ângela por ser a primeira a me apresentar ao mundo dos RNAs não codificadores e me acompanhar na minha iniciação científica e pela amizade.

À Lauren por sempre estar ao meu lado, pela cumplicidade e amizade. Muito obrigada por sempre me deixar dormir na sua casa.

Ao Santito pela amizade, por me apresentar o Fernet e abrir novos caminhos para minha vida.

À Katia por sempre discutir a maioria dos resultados comigo, por ter me dado grandes ideias e nunca ter me abandonado quando me sentia perdida. Muito obrigada amiga.

À Letícia por tornar os dias menos chatos com as suas músicas bregas e à Mariana por sempre dançá-las.

Ao Carlos por sempre ser sincero nos comentários.

Ao Yuri pela amizade e por me introduzir no mundo R. Obrigada por sempre acreditar em mim.

Ao Vinícius pela discussão e ensinamentos no mundo da bioinformática.

Ao Rodrigo Borges por sempre ter muita paciência comigo e me ensinar Perl.

Ao Murilo pelas danças de forró.

Ao Felipe e Ana Ayupe pelo apoio nos experimentos de *Northern blot* e pela amizade.

À Ana Paula pela amizade e por tornar a minha vida mais fácil. E ao Jefferson pelos cafés.

Ao Helder pelas festas e à Camila por poder sempre contar com ela mesmo de longe.

Aos outros colegas de trabalho: Otto, Kléber, Dinar, Giulliana, Gabriel e Ester.

À Dona Brega por sempre estar ao meu lado para o que der e vier e por ser minha irmã de verdade.

À Joana, que me faz muito falta, por ter feito parte de tudo isso, por ter me aguentado quando estava triste, brava e perdida. Obrigada pela amizade e companheirismo.

À Lu, por ser a pessoa mais centrada e sensata que eu conheço. Obrigada pela amizade e pelos valiosos conselhos.

Às meninas Maiara, Amira e Bahr, muito obrigada por me deixarem fazer parte da vida de vocês e por me distraírem com outros assuntos, deixando a minha vida mais legal.

À Evelyn e à Ju pelas conversas a tarde nos bancos do IQ, pelos risos, pela diversão e bebidas de graça. Adoro vocês.

À Paula pela companhia à tarde e churrascos na sua casa.

Ao José Geraldo por estar ao meu lado durante grande parte deste desafio. Pelo companherismo e amizade.

Ao Sofá pela amizade, apesar distante ainda continua a mesma diversão.

À minha família por sempre estar comigo e pelo apoio a esta dura jornada.

Aos meus primos pela amizade, pelas conversas e cumplicidade.

Ao Leandro por me apoior e acreditar bastante em mim.

Às meninas e aos meninos do time de futebol por manterem a minha mente ocupada com outras coisas.

Aos meus outros amigos que fizeram parte da vida.

À Fundação de Ampara à Pesquisa do Estado de São Paulo (FAPESP) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro.

A todos os funcionários do Instituto de Química pelos serviços prestados.

“O místico crê em um Deus desconhecido. O pensador e o cientista creem numa ordem desconhecida. É difícil dizer qual deles sobrepuja o outro em sua devoção não racional”

L. L. Whyte

“(...) Mas que preguiça boa me deixa aqui à toa. Hoje ninguém vai estraguar o meu dia.”

Charlie Brown Jr.

RESUMO

Tahira, A. C. **Análise da expressão de RNAs não codificadores em adenocarcinoma de pâncreas.** 2013. 207p. Tese de Doutorado – Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

RNAs não codificadores longos (lncRNAs) compõem uma fração significativa do transcriptoma. Alterações na expressão de lncRNAs já foram observadas em vários cânceres humanos, mas ainda não foram exploradas no adenocarcinoma pancreático ductal (PDAC), uma doença devastadora e agressiva para a qual faltam métodos para diagnóstico precoce e tratamentos efetivos. Utilizando uma plataforma de microarranjo de cDNA com sondas para 984 lncRNAs e 2371 mRNAs, o presente estudo identificou conjuntos de lncRNAs expressos em 38 amostras clínicas pancreáticas. O enriquecimento de (i) elementos regulatórios associados às regiões promotoras (H3K4me3); (ii) possíveis inícios de transcrição (*CAGE-tags*); (iii) presença de elementos conservados sugere que ao menos uma fração desses RNAs seja originada a partir de unidades transcricionais independentes, reguladas e possivelmente funcionais. Foram identificadas assinaturas de expressão gênica compostas por mRNA e lncRNAs associadas ao tumor primário e à metástase pancreática. A assinatura gIenica associada à metástase apresentou enriquecimento RNAs intrônicos de *loci* gênicos associados à via MAPK quinase. O aumento de expressão dos transcritos intrônicos dos *loci PPP3CB*, *MAP3K14* e *DAPK1* foi confirmado por qPCR em metástases. Em conjunto, este trabalho aponta para a importância de lncRNAs intrônicos no PDAC e para a necessidade de estudos mais aprofundados para uma melhor compreensão do papel dessa classe de transcritos na biologia da doença.

Palavras-chave: RNAs não codificadores longos, adenocarcinoma pancreático ductal, expressão gênica e microarranjos de cDNA.

ABSTRACT

Tahira, A. C. **Expression analysis of** l**ong noncoding RNAs in pancreatic adecarcinoma.** 2013. 207p. PhD Thesis – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Long noncoding RNAs (lncRNAs) compose a significant fraction of transcriptome. Altered expression of lncRNAs has been observed in diverse human cancers, but has not being investigated in pancreatic ductal adenocarcinoma (PDAC), a devastating and aggressive disease that lack early diagnosis methods and effective treatments. Using a cDNA microarray platform with probes interrogating 984 lncRNAs and 2371 mRNA, the present study identified subsets of lncRNAs expressed in 38 pancreatic clinical samples. Enrichment of (i) regulatory elements associated to promoter region (H3K4me3); (ii) putative transcription start site (CAGE*tags*) and (iii) conserved elements, suggest that at least a fraction of these RNAs could be independent transcriptional unit, regulated, an possibly functional. Gene expression signatures comprised of mRNAs and lncRNAs and associated to primary or metastatic tumors were found. A gene signature associated to metastasis was enriched in intronic ncRNAs mapping to gene *loci* associated to the MAPK pathway. Over expression of intronic RNAs from *PPP3CB*, *MAP3K14* and *DAPK1* was confirmed by qPCR in metastatic samples. Taken together, this study points to the importance of intronic lncRNAs in PDAC and for the need to study this class of ncRNAs in greater detail to better understand its role in the biology of PDAC.

Keywords: long noncoding RNAs, pancreatic ductal adenocarcinoma, gene expression and cDNA microarray.

LISTA DE ABREVIATURAS E SIGLAS

aa: aminoácidos

cDNA: DNA complementar

CPC: *Coding Potencial Calculator*

cRNA: RNA complementar

ENCODE: *Encyclopedia of DNA Elements*

EST: *Expressed Sequence Tags*

EZH2: *Enhancer of zeste homolog 2*

FANTOM: *Functional Annotation of Mouse Consortium*

FDR: *False Discovery Rate*

H3K4me3: trimetilação da lisina 4 da histona 3

H3K36me3: trimetilação da lisina 36 da histona 3

IPMN: neoplasia intraductal mucinosa papilífera

lncRNAs: *long noncoding RNA*

lincRNA: *long intergenic RNA*

MCN: cistoadenocarcinoma mucinoso

mRNA: RNA mensageiro

miRNA: microRNAs

MPSS: *Massively Parallel Signature Sequencing*

ncRNA: *noncoding RNA*

PanIN: neoplasia intraepitelial pancreática

PASR: *promoter-associated short RNA*

PDAC: *Pancreatic ductal adenocarcinoma*

PCR: *Polymerase Chain Reaction*

PCR2: *Polycomb Repressor Complex 2*

PIN: *Patially Intronic transcripts*

qPCR: PCR quantitativo

RIN: *RNA Integrity*

RPKM: *Reads per kilobase per million*

SAGE: *Serial Analysis of Gene Expression*

SAM: *Significance Analysis of Microarray*

snoRNA: *small nucleolar RNA*

SNP: *Single nucleotide polymorphins*

snRNA: *small nuclear RNA*

TIN : *totally intronic transcript*

tiRNA: *tiny RNA*

T-UCR: *transcribed ultraconserved region*

UTR: *untranslated region*

SUMÁRIO

1 INTRODUÇÃO 16
1.1 O câncer de pâncreas 16
1.1.1 Lesões precursoras no adenocarcinoma pancreático. 17
1.1.2 Pancreatite crônica e o PDAC. 22
1.1.3 Marcadores moleculares no PDAC 23
1.1.4 Uso de metodologias de larga escala para identificação de biomarcadores no PDAC 25
1.2 A complexidade do genoma humano e RNAs não codificadores (ncRNAs). 27
1.3 LncRNAs e câncer 33
2 OBJETIVOS 36
2.1 Objetivos gerais 36
2.2 Objetivos específicos 36
3 MATERIAIS E MÉTODOS 37
3.1 Coleta e armazenamento de amostras 37
3.2 Extração de RNA das amostras clínicas de pacientes 39
3.3 Avaliação da contaminação genômica (DNA) no RNA total 43
3.4 Amplificação linear no RNA mensageiro (mRNA) 44
3.4.1 Purificação do cDNA dupla-fita 48
3.4.2 Marcação de alvos fluorescentes a partir de cRNA com Cy5 49
3.5 Hibridização de cRNA com microarranjos de DNA 50
3.6 Análise informática dos dados de expressão gênica 52

3.7 Análise bioinformática dos transcritos intrônicos e intergênicos expressos nos tecidos pancreáticos. 56

3.7.1 Comparação do tamanho das ESTs utilizadas como sondas no microarranjo com a sequencia de ncRNAs preditos a partir de uma reconstrução bioinformática do transcriptoma. 57

3.7.2 Avaliação da expressão dos ncRNAs interrogados no microarranjo em diferentes tecidos utilizando dados de RNA-Seq disponíveis publicamente. 58

3.7.3 Avaliação da localização de ncRNAs intergênicos interrogados no microarranjo relativamente a regiões 3' ou 5' UTR de RNAs codificadores de proteínas conhecidos . 59

3.7.4 Procura por elementos conservados no DNA nas sequências dos RNAs expressos em tecidos pancreáticos 60

3.7.5 Predição de estrutura secundária e potencial codificador 61

3.7.6 Transmap – análise de conservação de expressão 63

3.7.7 Identificação de elementos regulatórios na vizinhança de ncRNAs 64

3.8 Enriquecimento de categorias associadas a Gene Ontology (GO) 65

3.9 RT-PCR fita-específica 67

3.10 RT-PCR quantitativo em tempo real 68

3.11 Clonagem de ncRNAs através de RACE. 69

3.12 *Northern Blot*. 70

3.13 Ensaio de promotor in vitro utilizando um gene repórter 71

3.13.1 Busca por predição de regiões promotoras *in silico* 71

3.13.2 Clonagem das possíveis regiões promotoras no vetor pGL3 (Promega) 72

3.14 Transfecção dos vetores pGL3 em células Mia PaCa-2 76

3.15 Cultivo de Células Mia PaCa-2, PANC-1, AsPC-1 e Capan-1..... 77

3.16 Silenciamento de transcritos utilizando oligos antisenso modificados..... 78

4 RESULTADOS 80

4.1 RNAs não codificadores longos intrônicos e intergênicos são expressos em tecidos pancreáticos com diferentes graus de malignidade..... 80

4.2 Análise *in silico* do potencial codificador, tamanho e processamento em pequenos RNAs dos transcritos intrônicos expressos em tecidos pancreáticos..... 87

4.3 Busca por elementos regulatórios associados às regiões promotoras (H3K4me3 e CpG) e evidências de 5'-CAP nos transcritos expressos em tecidos pancreáticos..... 96

4.4 Identificação de assinaturas de expressão gênica de mRNAs e ncRNAs (intrônicos e intergênicos) correlacionadas com adenocarcinoma pancreático. 100

4.5 Identificação de genes correlacionados com a metástase do câncer de pâncreas. ... 112

4.6 LncRNAs intrônicos associados à metástase pancreática são expressos na orientação senso ou antissenso em relação ao mRNA (codificador de proteína)..... 120

4.7 LncRNAs associados à metástase pancreática apresentam tamanho médio de aproximadamente 1kb. 124

4.8 LncRNAs senso *DAPK1* possui tamanho de 1050pb confirmados por RT-PCR fita-específica em MIA PaCa-2 e é expresso em Capan-1. 129

4.9 LncRNA senso *DAPK1* possui região promotora e pode estar associado a presença das proteínas de isoforma curta e longa. 131

4.10 A expressão do lncRNAs senso *DAPK1* não parece estar correlacionada com as isoformas de mRNAs dos correspondentes codificadores de proteína do *locus DAPK-1*.. 134

4.11 A expressão do lncRNA senso DAPK1 parece estar correlacionada com a dapk1 isoforma longa em linhagens pancreáticas MIA PaCa 2 e Capan-1. 137

5 DISCUSSÃO 142

6 CONCLUSÕES 162

7 REFERÊNCIAS 164

APÊNDICE 177

ANEXOS ELETRÔNICOS 208

ANEXO 209

1 Introdução

1.1 O câncer de pâncreas

O adenocarcinoma pancreático ductal (PDAC) é o tipo mais comum de neoplasia pancreática representando mais de 85% dos casos de câncer de pâncreas (Hezel et al. 2006). O PDAC é uma doença devastadora que apresenta mau prognóstico (Yokoyama et al. 2009) e sua agressividade está associada à falta de um diagnóstico precoce e à resposta limitada aos tratamentos disponíveis, contribuindo para alta taxa de mortalidade (Gironella et al. 2007). Apesar pacientes que se submeteram à ressecção cirúrgica terem uma taxa de sobrevida de 5 anos de 20%, a maioria dos pacientes não é candidato à cirurgia, pois a doença é diagnosticada em estágios avançados e, na maior parte dos casos, já sofreu metástase (Greer et al. 2007). Desta forma, apenas uma pequena parte dos pacientes (15-20%) é indicada para este procedimento (Siegel et al. 2011). Da parcela que é indicada para resecção cirúrgica, apenas 20% possuem taxa de sobrevida de 5 anos, resultando em uma taxa de sobrevida de 5 anos total de 3-5% (Hezel et al. 2006). Além da dificuldade em realizar um diagnóstico precoce, a agressividade do adenocarcinoma pancreático está intimamente associada à resposta limitada aos tratamentos disponíveis (Yokoyama et al. 2009). Mesmo com combinações altamente tóxicas de quimioterápicos, a sobrevida média do paciente é menor do que um ano (Conroy et al. 2011). Muitas drogas já testados mostraram bons resultados em ensaios *in vitro* com linhagens celulares pancreáticas e *in vivo* com animais. Entretanto, até o momento nenhuma das drogas testadas em ensaios clínicos resultou em benefício para pacientes com câncer de pâncreas (Paez-Ribes et al. 2009; Kindler et al. 2010; Conroy et al. 2011; Kindler et al. 2012). Por exemplo, o estudo conduzido por Moore e colaboradores

mostrou apenas um discreto aumento na taxa de sobrevida com o uso de *erlotinib*[1] com *gemcitabine*[2], sendo de 5,91 meses (*gemcitabine* e placebo) para 6,24 meses (*erlotinib* e *gemcitabine*) em pacientes com câncer de pâncreas avançado (Moore et al. 2007). Mais recentemente, o estudo conduzido por Kindler e colaboradores, o uso de *sorafenib*[3] em combinação com *gemcitabine* em pacientes com câncer de pâncreas avançado não trouxe benefícios clínicos para os pacientes (Kindler et al. 2012).

1.1.1 Lesões precursoras no adenocarcinoma pancreático.

Após duas décadas de estudos moleculares associados ao adenocarcinoma pancreático, acredita-se que esta seja uma doença genética e progressiva e muitas das mutações são recorrentemente encontradas em câncer de pâncreas (Maitra and Hruban 2008). Em um estudo conduzido por Jones e colaboradores mostrou a presença de 63 alterações genéticas recorrentes no PDAC, na maioria mutações pontuais, que afetam 12 vias de sinalização celular principais, onde cada via estava alterada numa taxa entre 67 a 100% nos 24 tumores pancreaáticos em estágios avançados analisados (Jones et al. 2008). Além disso, muitas das mutações encontradas no câncer de pâncreas são encontradas na lesões pré-malignas (Singh and Maitra 2007). As lesões precursoras que podem dar origem ao PDAC são conhecidas como *PanIN*[4] (neoplasia intraepitelial pancreática), *IPMNs* (neoplasia intraductal mucinosa papilífera) e *MCNs* (cistoadenocarcinoma mucinoso) (Figura 1).

---

[1] *Erlotinib*: inibidor de tirosina quinase que age no receptor de fator de crescimento epidérmico tipo 1 (EGFR/HER1).
[2] *Gemcitabine*: análogo de nucleosídeo, no qual o átomo de hidrogênio localizado no 2'C da deoxicitidina é substituído por átomos de flúor.
[3] *Sorafenib*: pequena molécula inibidora de diversas tirosinas e *Raf* quinases (*B-raf*, *VEGFR2* e *PDGFR-beta*).
[4] *PanINs*: representa lesões precursoras de pequeno calibre (<5 mm) originadas do ducto pancreático (Hruban et al. 2000)

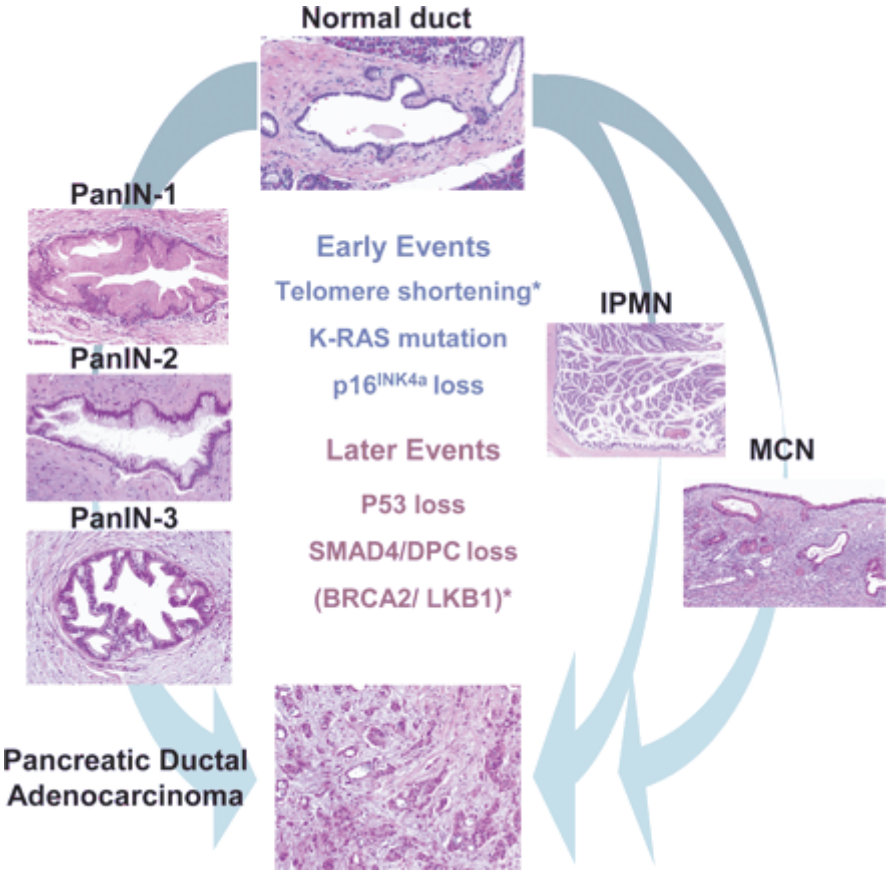

Figura 1: Lesões pancreáticas precursoras e eventos genéticos envolvidos na progressão do PDAC. A figura acima representa as 3 lesões precursoras do PDAC: *PanIN*, *MCN* e *IPMN*. A progressão da lesão *PanIN* está mostrada à esquerda, o aumento do grau que varia de 1 a 3 reflete no aumento de atipia celular e, eventualmente leva à formação de PDAC. Ao lado direito está representada a progressão potencial das lesões *MCNs* e *IPMNs* para a formação de PDAC. As alterações genéticas mostradas acima ocorrem em lesões *PanIN*, e numa menor extensão em *MCNs* e *IPMNs*, em uma aparente sequência temporal de eventos, embora essas alterações não terem sido correlacionadas com aquisição específica referentes às características histopatológicas. Os vários eventos genéticos estão listados e divididos entre, predominantemente, aqueles que ocorrem em estagios iniciais e tardios da progressão do PDAC. Os asteriscos representam eventos que não são comuns a todas as lesões precursoras (o encurtamento dos telômeros e perda de *BRAC2* são relatados em *PanIN* e perda de *LKB1*é documentada em subconjuntos de PDAC e *IPMNs*). Figura retirada de Hezel (Hezel et al. 2006).

Apesar do adenocarcinoma poder se originar de qualquer uma das lesões pré-malignas acima, a mais comumente encontrada é a *PanIN*. A lesão pré-maligna *PanIN* é conhecida pela existência de uma progressão no estágio da lesão caracterizando pelo aumento de atipia de

células do ducto, denominadas *PanIN-1*, *PanIN-2* e *PanIN-3* (Iacobuzio-Donahue 2012). O encurtamento dos telômeros em comparação tecido normal é encontrado em todos os estágios das lesões pré-malignas *PanIN* (van Heek et al. 2002). A mutação no gene *KRAS2*[5], o qual resulta na ativação constitutiva da atividade GTPase intrínseca da proteína, independente da transdução de sinal intra ou extracelular (Schubbert et al. 2007), já é detectada em lesões precursoras *PanIN-1A* (Moskaluk et al. 1997). As lesões *PanIN-2* apresentam a perda de função de *CDKN2A*[6], que é vista em quase 90% dos casos de adenocarcinoma pancreático. Esta perda pode ocorrer através de vários mecanismos como deleção homozigótica (40%), mutação intragênica com perda do segundo alelo (40%) ou diminuição da expressão por silenciamento epigenético por meio da metilação da região promotora (10-15%) (Caldas et al. 1994; Schutte et al. 1997; Ueki et al. 2000). Observa-se com frequência a inativação de *p53*[7] e *SMAD4*[8] em lesões de *PanIN-3*, a qual está presente em 50-75% e 55% dos cânceres de pâncreas, respectivamente (Redston et al. 1994; Hahn et al. 1996). A inativação de *TP53* ocorre quase sempre pela mutação intragênica, que leva a perda do segundo alelo (Vogelstein et al. 2000). Já a perda de *SMAD4* pode ocorrer de duas principais maneiras; por deleção homozigótica (30%) ou por mutação intragênica com perda do segundo alelo (25%) (Hahn et al. 1996).

---

[5] *KRAS2*: gene que faz parte da família RAS que se liga à proteína GTP, a qual medeia diversos processos celulares incluindo proliferação, sobrevivência e remodelamento do citoesqueleto.

[6] *CDKN2A*: também conhecido com *p16* e *INK4A*, este gene faz parte da família de *CDK* (ciclinas dependentes de quinase) e possui o papel de controlar a progressão do ciclo celular através do ponto de parada G1-S mediada por *CDK4* e *CDK6* (Kim and Sharpless 2006)

[7] *p53*: gene localizado no cromossomo 17, sendo um regulador importante para vários diversos processos celulares, incluindo regulador de ponto de checagem G1-S, parada do ciclo celular G2-M e indução de apoptose após estímulos de *stress* e dano do DNA (Raycroft et al. 1990)

[8] *SMAD4*: regulador importante da via de sinalização de TGF-β responsável pela fosforilação de fatores de transcrição Smad 2/3, que depois formam complexos (Smad 4/2/3) que entram no núcleo para ativar ou reprimir a expressão gênica.

As lesões pré-malignas *IPMNs* são neoplasias epiteliais que ocorrem no ducto pancreático principal ou em suas ramificações (Belyaev et al. 2008). Em geral, este tipo de lesão afeta mais homens do que mulheres em uma taxa de (3:2) e é frequentemente encontrada na cabeça do órgão pancreático(Salvia et al. 2004). Um terço das lesões deste tipo estão associadas ao carcinoma invasivo (Matthaei et al. 2011). A importância clínica de detectar este tipo de lesão em estágios iniciais foi demonstrada por estudos, os quais mostraram que pacientes que apresentavam *IPMNs* não invasivas e foram submetidos à resecção cirúrgica apresentaram taxa de sobrevida de 5 anos acima de 90%, enquanto pacientes que apresentavam *IPMNs* invasivo apresentavam aproximadamente 50% de taxa de sobrevida de 5 anos (Salvia et al. 2004; Sohn et al. 2004). Na revisão publicada por Sugiyama e colaboradores (Sugiyama et al. 2008), a taxa de sobrevida de 5 anos de pacientes com lesões benignas chegavam próximo aos 100%, enquanto que para aqueles que apresentavam lesões malignas a taxa variava entre 24-74% (média de 60%). Estudo publicado recentemente por Yamaguchi e colaboradores (Yamaguchi et al. 2011) mostrou que PDAC originados de lesões IPMN são menos agressivos que os adenocarcinomas pancreáticos ductais ordinários, enquanto o primeiro apresenta uma taxa de sobrevida em torno de 47 meses. Os sintomas apresentados por esse tipo de lesão pancreática são muito comuns e inespecíficos (dor abdominais e nas costas, anorexia, perda de peso e recorrente episódios de pancreatite), representando uma grande dificuldade de diagnóstico precoce na ausência de biomarcadores confiáveis (Matthaei et al. 2011).

Similar às lesões *IPMNs*, aproximadamente, um terço das lesões *MCNs* estão associadas ao adenocarcinoma invasivo do tipo ductal (Crippa et al. 2010). Este tipo de lesão pancreática é a mais rara dentre as três apresentadas e afeta muito mais mulheres do que homens numa taxa de 20:1. As alterações moleculares são muito parecidas com as apresentadas por *PanIN*, logo, a ativação do gene *KRAS* é observada mesmo em displasia de

baixo grau, enquanto mutações em *TP53* e *SMAD4* são encontradas em displasias mais avançadas (Jimenez et al. 1999; Iacobuzio-Donahue et al. 2000). Além disso, a expressão *MUC5A* e *MUC* estão associadas com características não invasivas de *MCNs*, enquanto que a expressão de *MUC1* está associada à propriedades invasivas (Luttges et al. 2002).

Uma das características do câncer de pâncreas é sua disseminação sistemática e extraordinária progressão local (DiMagno et al. 1999). O trabalho publicado recentemente por Rhim e colaboradores (Rhim et al. 2012), identificou que a invasão de células pancreáticas em outros órgãos (ex. fígado) ocorre em estágios que se acredita serem ainda pré-invasivos. Este estudo sugere que a invasão de outros órgãos ocorra antes, ou em paralelo à formação do tumor no sítio primário. Esta evidência ainda é suportada pelo fato de que 75% de pacientes submetidos à resecção cirúrgica de pequenos tumores pancreáticos com margens livres e nenhuma evidência de metástase acabam morrendo de metástase pancreática dentro de 5 anos (Neoptolemos et al. 2004). Neste mesmo trabalho foi visto que a presença de inflamação promove a transição epilélio-mesenquimal e disseminação por dois mecanismos distintos: mudanças do microambiente no sítio primário e maior acesso à circulação sanguínea da célula (Rhim et al. 2012).

A inflamação está comumente associada com iniciação do tumor e progressão (Grivennikov et al. 2010) e foi visto no trabalho de Guerra e colaboradores (Kreutzer et al. 2010) que quando o oncogene *Kras* (G12V) foi especificamente expresso em camundongos adultos, nenhuma transformação maligna (ou pré-maligna) ocorreu, entretanto após o estímulo inflamatório, houve a formação de câncer pancreático em camundongos *Kras* mutantes adultos (Guerra et al. 2007). Este trabalho suporta a ideia de que a inflamação crônica do estroma é, possivelmente, um evento adicional que ativa e perpetua a transformação maligna do adenocarcinoma pancreático (Erkan et al. 2012). Além disso, já foi

visto que tumores pancreáticos mucinosos são comumente associados a uma pequena presença de células estromais ao redor do tumor, o que também é acompanhado por um fenótipo menos agressivo quando comparado aos tumores com uma alta presença de células estromais (Kimura et al. 1998). Estas evidências levam a crer que a presença de células estromais possui uma contribuição ativa no processo de carcinogênese (Kong et al. 2012).

1.1.2 Pancreatite crônica e o PDAC.

A pancreatite crônica é resultado do processo inflmatório no pâncreas, sendo uma doença debilitante que causa dores abdominais e mudanças morfológicas e funcionais no pâncreas, juntamente, com uma queda na qualidade de vida do paciente (Yin et al. 2012). Os principais fatores associados ao aparecimento desta doença são o consumo de cigarro e álcool (Talamini et al. 1999; Lowenfels et al. 2001). Já foi mostrado que pacientes que apresentam pancreatite crônica possuem maior risco de desenvolver adenocarcinoma pancreático (Yin et al. 2012). Além disso, pessoas que possuem pancreatite autoimune, uma doença genética autossômica dominante, também apresentam maior risco de desenvolver câncer de pâncreas (Lowenfels et al. 1997). Uma das explicações para isto é o processo inflamatório causado pela doença, e a sua reincidência é um fator indireto para o aumento do risco do desenvolvimento de PDAC (Whitcomb 2004; Dite et al. 2012). O processo inflamatório contribui para o aparecimento do câncer de pâncreas, pois vem acompanhado de uma reação demosplásica caracterizada pelo aumento de componentes estromais e ativação de PaSCs (*Pancreatic Stellate Cells*), as quais são responsáveis pela a deposição de ECM (*Extra cellular matrix*) (Feig et al. 2012). Suspeita-se que o acúmulo dessa matriz comprime as veias sanguíneas, diminuindo o acesso de drogas antitumorais às células neoplásicas (Feig et al. 2012), uma vez que já foi visto em camundongos que o câncer de pâncreas é hipovascularizado (Olive et al. 2009). Este processo inflamatório encontrado em pacientes com adenocarcinoma pancreático

é muito parecido com o visto em pacientes com pancreatite crônica (Mahadevan and Von Hoff 2007). Desta forma, alguns estudos de expressão gênica relacionados ao adenocarcinoma pancreático utilizam amostras de pancreatite crônica para identificar quais seriam genes relacionados à presença de células estromais (Logsdon et al. 2003; Bloomston et al. 2007). Uma alternativa para atenuar a influência das células estromais em estudos de biomarcadores, muitos grupos utilizam a microdissecção a laser nas amostras de pacientes, enriquecendo-as com células tumorais ou não tumorais (Crnogorac-Jurcevic et al. 2003; Grutzmann et al. 2004; Grutzmann et al. 2005; Badea et al. 2008; Lopez-Casas and Lopez-Fernandez 2010).

1.1.3 Marcadores moleculares no PDAC

Apesar de o adenocarcinoma pancreático ser caracterizado como uma doença genética progressiva como descrito por Campbell e colaboradores (Campbell et al. 2010) e Yachida e colaboradores (Yachida et al. 2010), o uso de marcadores moleculares ou novos alvos terapêuticos ainda é muito escasso. O trabalho de Yachida e colaboradores apresenta a teoria de que a metástase ocorre em estágios tardios, demorando mais de uma década entre o aparecimento da primeira célula tumoral até o aparecimento de metástase, sugerindo que há uma grande janela de oportunidade para o diagnóstico precoce do adenocarcinoma pancreático (Yachida et al. 2010). Entretanto, a falta de sintomas nos pacientes que possuem lesões pré-malignas e a localização anotômica do pâncreas dificultam a detecção precoce da doença (Mohammed et al. 2012). Apesar de já serem conhecidas diversas alterações genéticas associadas ao PDAC, ainda não existem, de fato, biomarcadores confiáveis para o diagnóstico precoce e prognóstico da doença (Fong and Winter 2012; Winter et al. 2012). O marcador mais comumente utilizado para indicar progressão e recorrência da doença é o

CA19-9[9] sorológico (Pleskow et al. 1989). Apesar do aumento do nível de *CA19-9* ser útil em diferenciar entre adenocarcinomas pancreáticos e condições inflamatórias do pâncreas, a sensibilidade (70-80%) e especificidade (68%-91%) apresentadas não são consideradas suficientes (Passerini et al. 2012). Outros tipos de biomarcadores já foram investigados no soro como CA125 e CEA para detectar adenocarcinoma pancreático, entretanto, foram menos informativos que o *CA 19-9* (Winter et al. 2012). Esforços para identificar novos biomarcadores têm sido realizados como o trabalho de Harsha e colaboradores que montou um catálogo com todos os possíveis biomarcadores investigados no adenocarcinoma (Harsha et al. 2009). Este compedio teve como base vários trabalhos de larga-escala utilizando microarranjos e outras evidências experimentais. Ainda, atualmente, existem bancos de dados disponíveis sobre o câncer de pâncreas como *Pancreatic Expression Database* (http://pancreasexpression.org/) e *Pancreatic Cancer Database* (http://pancreaticcancerdatabase.org/), os quais possuem trabalhos da literatura baseados em técnicas de análise em larga-escala referentes ao genoma, transcriptoma e proteômica, além de outras evidências experimentais (qPCR, *western/northern blot* entre outras abordagens). Um dos exemplos catalogado no compendio e nesses bancos de dados é a proteína S100A6[10], a qual já foi associada à baixa taxa de sobrevida de pacientes com adenocarcinoma pancreático que apresentavam marcação nuclear nas células malignas (Vimalachandran et al. 2005). Este gene também foi apontado como um candidato para biomarcador discriminar pancreatite crônica e PDAC, entretanto não foi capaz de discriminar entre PDAC e IPMN (Ohuchida et al. 2007). As proteínas TIMP1, MMP9 e MMP11, identificadas como aumentadas em adenocarcinomas pancreáticos (Harsha et al. 2009) também já foram investigadas como biomarcador. Ainda, a proteína S100P, da família de proteínas ligadoras de

---

[9] *CA19-9*: *carbohydrate antigen 19-9*, também conhecido como *cancer antigen 19-9* ou *sialylated Lewis (a) antigen*, é um carboidrato produzido pelas células epiteliais exócrinas e normalmente, absorvida pela superfície dos eritrócitos.
[10] *S100A6*: *S100 calcium binding protein A6.*

cálcio, foi mostrada expressa exclusivamente em células cancerosas do pâncreas (Dowen et al. 2005).

Recentemente, o estudo de pequenos RNAs não codificadores (ncRNAs) com 21-22nt denominados microRNAs têm chamado a atenção por estarem envolvidos em diversos processos biológicos inclusive câncer (Ebert and Sharp 2012; Lages et al. 2012). Vários estudos realizados identificaram no conjunto mais de 20 miRNAs envolvidos na patologia do adenocarcinoma pancreático, apontando os miRNAs como moléculas promissoras para o diagnóstico e como ferramenta terapêutica no tratamento de adenocarcinoma pancreático (Papaconstantinou et al. 2012). Assim, é necessária uma maior compreensão das bases moleculares associadas ao câncer de pâncreas de forma a permitir a identificação de novos e/ou mais eficientes marcadores para diagnóstico e alvos terapêuticos desta doença.

1.1.4 Uso de metodologias de larga escala para identificação de biomarcadores no PDAC

Metodologias de larga escala baseadas em plataforma de microarranjos para estudar padrão de expressão gênica tem sido usadas extensivamente para a busca de genes diferencialmente expressos em câncer (Virtanen and Woodgett 2008). Esta técnica se baseia na hibridação complementar entre nucleotídeos e pode ser aplicada para estudar diversos genes simuntaneamente (Khan et al. 1999). Desde a publicação de Gress e colaboradores (Gress et al. 1996), a qual identificou 369 ESTs diferencialmente expressas no adenocarcinoma pancreático, o uso de plataformas de microarranjos tem sido aplicado em vários estudos realcionados ao PDAC (Buchholz et al. 2001; Iacobuzio-Donahue et al. 2003; Grutzmann et al. 2005; Badea et al. 2008; Buchholz et al. 2008; Cao et al. 2008; Stratford et al. 2010; Sergeant et al. 2012). Além do uso de plataformas de microarranjos, o uso de outras técnicas em larga escala tem sido aplicado para identificar mutações como mostrado no trabalho pioneiro de Jones e colaboradores (Jones et al. 2008), o qual investigou mutações em

20.735 genes em 24 amostras de pacientes com PDAC. Esta análise apontou para 69 genes alterados na maioria das 24 amostras, sendo 31 desses agrupados em 12 vias, que estavam alteradas em 67-100% das amostras analisadas. A fim de verificar se essas mutações refletiam na expressão aberrante desses genes, foi realizada análise de transcriptoma utilizando *SAGE*. O conjunto de 31 genes mutados possuía um maior enriquecimento de genes diferencialmente expressos em comparação aos 3041 genes que foram analisados, mostrando, que essas 12 vias possam ser importantes para o PDAC.

Com o avanço tecnológico, atualmente, há um apelo para o uso de tecnologias de nova geração para o estudo molecular de doenças como o adenocarcinoma pancreático (Mardis 2012). Estudo publicado por Campbell (Campbell et al. 2010) analisou 13 PDAC, dos quais 3 eram provenientes de tumores primários e 10 de metástases em múltiplos sítios. Neste trabalho, os rearranjos cromossômicos encontrado em amostras metastáticas foram comparados com o tumor primário pareado. Foi visto que muitos arranjos cromossômicos presentes na metástase não estavam representados nas amostras de tumores primários, sugerindo que estes rearranjos cromossômicos possam ser importantes para o processo metastático. Contudo, ao comparar os rearranjos presentes nas amostras metastáticas, foi visto que as 10 amostras analisadas mostravam rearranjos cromossômicos, na maior parte, diferente entre si. Esta heterogeneidade de mutações entre amostras de pacientes pode ser vista no trabalho pulibado por Yachida e colaboradores (Yachida et al. 2010), onde 7 amostras de pacientes foram analisadas quanto a taxa de mutação (SNPs). Muitas mutações, a qual o autor denominava fundadora, eram provenientes de genes diferentes entre as amostras de pacientes e também entre amostras do mesmo paciente, mostrando que além da heterogneidade inter amostras, há também uma heterogeneidade na população de células que compõem o tumor do prórpio paciente. Ainda, no trabalho realizado por Biankin e colaboradores (Biankin et al. 2012), o qual sequenciou o exoma de 99 amostras clínicas de paciente, identificou mutações

em genes que estão associados à via de *axon guidance* (*ROBO1*, *ROBO2*, *SLIT2*, *SEMA3A*, *SEMA3E*, *SEMA5A*, *EPHA5* e *EPH7*), apesar do gene *ROBO2* apresentar uma taxa de mutação em apenas 5% dos pacientes analisados, a expressão do mRNA foi capaz de discernir entre pacientes de bom e mau prognóstico. Interessantemente, o estudo publicado recentemente por Wang e colaboradores (Wang et al. 2012), o qual realizou sequenciamento do exoma e do transcriptoma de 15 linhagens celulares, também identificou uma grande heterogeidade entre as mutações gênicas encontradas nas linhagens celulares, entretanto fez uma correlação com a taxa de mutações *indels* (inserções e deleções) com a presença ou ausência do gene *MLH1*, o qual é reponsável pelo reparo deste tipo de mutação, sugerindo que a ausência deste gene possa estar contribuindo para instabilidade genômica e o aparecimento do PDAC. Análises de RNA-seq em *CTC* (*circulationg tumor cells*) em camudongos modelos para adenocarcinoma pancreático apontaram para a expressão do gene *WTN2* e, isto foi confirmado em 5 amostras clínicas de pacientes das 11 analisadas. Apesar do avanço tecnológico, ainda são utilizadas plataformas de microarranjos para avaliar genes com possíveis marcadores no adenocarcinoma pancreático (Stratford et al. 2010; Sergeant et al. 2012), por apresentarem baixo custo e pouca manipulação da amostra (Baginsky et al. 2010).

1.2 A complexidade do genoma humano e RNAs não codificadores (ncRNAs).

O dogma central da biologia molecular correspondente ao fluxo de informação genética descreveu o mRNA como a molécula intermediária que levaria a informação para a síntese de proteína proposto por Francis Crick em 1958 (Crick 1958) e revisado em 1970 (Crick 1970). Pelo fato das proteínas desempenharem papéis cruciais em todos os processos celulares, acreditava-se que a complexidade de organismo estava relacionada com a quantidade de proteínas, e, portanto, com os genes que a codificavam. Assim, estimativas para o número de

genes codificadores de proteína em humanos chegaram a 35.000 a 140.000 (Board 2000; Ewing and Green 2000; Liang et al. 2000). Entretanto, com o sequenciamento do genoma humano em 2001 (Lander et al. 2001; Venter et al. 2001), foi observado que apenas uma pequena (1.5 - 2%) do genoma era composta por partes codificadoras de proteínas, e, que representavam menos de 25.000 genes codificadores de proteína distintos (Consortium 2004). Comparativamente, este número é muito similar ao encontrado em planta *Arabidopsis thaliana* (Initiative 2000) e não se apresenta muito maior do que em eucariotos inferiores, *Drosophila meloganaster* e *Caenorhabditis elegans*, que apresentam 13.600 (Adams et al. 2000) e 19.000 (Consortium 1998) genes codificadores, respectivamente. Desta forma, fica claro que a complexidade do organismo não deve estar somente correlacionada com o número de genes codificadores de proteína.

Diversas formas de regulação como o uso diferencial de isoformas de proteínas (*splicing* alternativo), controle transcricional e de tradução ou modificação pós-traducional de proteínas (Venter et al. 2001; Harrison et al. 2002) contribuem para aumentar a complexidade da expressão gênica dos organismos. Além disso, regiões do DNA compostas por RNAs não codificadores (ncRNAs) representam uma camada de regulação, a qual poderia direcionar a diferenciação e desenvolvimento de células e tecidos, por meio de uma série de mecanismos (*imprinting*, regulação de promotores, *splicing* alternativo, edição de RNA, estabilidade de mRNA e tradução de mRNA (Mattick 2003; Mattick 2004). Como uma evidência de que ncRNAs podem contribuir para o ganho de complexidade celular, tecidual e fenotípica, tem sido notado que organismos mais complexos, como eucariotos superiores, possuem proporção de regiões não codificadoras em relação ao tamanho do genoma maior do que organismos mais simples, como procariotos e eucariotos inferiores (Mattick 2004).

A primeira evidência de que ncRNAs poderiam exercer papel funcional na célula foi observada com a caracterização do RNA não codificador longo *XIST*[11]no começo dos anos 90 (Brown et al. 1991). Atualmente, sabe-se que este lncRNA (RNA não codificador longo) é transcrito a partir do cromossomo X inativo e recruta o complexo repressivo *Polycomb 2* (*PCR2*), o qual realiza a trimetilação da lisina 27 da histona 3 (H3K27me3) para silenciar a transcrição do *locus* (Nagano and Fraser 2011). Com a utilização extensiva de análises em larga escala do transcriptoma de eucariotos verificou-se que os genomas eucarióticos produzem uma grande variedade de isoformas alternativas de s*plicing* para quase todos os genes codificadores de proteína e ncRNAs,denominados *TUFs* (*transcripts of unknown function*), RNAs antissenso e uma grande quantidade de ncRNAs que compõe uma complexa rede transcricional e que parecem possuir funções biológicas importantes (Willingham and Gingeras 2006). Um trabalho pioneiro desenvolvido por Kapranov e colaboradores procurou identificar a atividade transcricional dos cromossomos 21 e 22 em 11 linhagens celulares diferentes utilizando uma plataforma de microarranjos de alta densidade (*tilling array*). Neste estudo foi observado que 94% das regiões transcritas em pelo menos uma das 11 células estudadas mapeava fora de regiões anotadas como exons de genes conhecidos (Kapranov et al. 2002). Abordando os mesmos cromossomos 21 e 22 e também utilizando *tilling arrays*, Kampa e colaboradores identificaram que apenas 31.4% dos nucleotídeos transcritos mapeavam em genes bem anotados, o restante possuía evidência de transcrição baseada em mRNAs ou ESTs (*Expressed Sequence Tags*) ou não possuía qualquer evidência de serem transcritos (Kampa et al. 2004). Ainda, neste mesmo trabalho foi visto que parte dos nucleotídeos transcritos em regiões exônicas produziam transcritos com orientação antissenso ao mRNA codificador de proteína. Transcritos antissenso de genes conhecidos também foram detectados no cromossomo 22 no trabalho de Rinn e colaboradores (Rinn et al. 2003).

[11] *XIST* (*X_i-specific transcripts*): transcrito de aproximadamente 17kb, que silencia um dos cromossomos X.

Consistente com as observações de que regiões não codificadoras do genoma são transcritas, outros trabalhos também apontaram para essas evidências utilizando microarranjos (*tilling*) que interrogavam regiões genômicas maiores (Bertone et al. 2004; Cheng et al. 2005; Kapranov et al. 2005; Kapranov et al. 2007), ou utilizando outras técnicas como SAGE (*Serial Analysis of Gene Expression*) (Chen et al. 2002; Saha et al. 2002), ou MPSS (*Massively Parallel Signature Sequencing*) (Jongeneel et al. 2005) ou sequenciamento de bibliotecas de cDNAs em camundongos (Carninci et al. 2005; Katayama et al. 2005).

Apesar da evidência de transcrição pervasiva do genoma, muitos grupos apontaram a possibilidade de que esses transcritos poderiam representar pré-mRNAs de genes codificadores de proteínas abundantes (van Bakel et al. 2010), ser produto de transcrição espúria catalisada pela RNA polimerase II (Struhl 2007) ou ainda representar introns processados do pré-mRNA denominados intron *lariats* (Cheng and Menees 2011). Contudo, um conjunto extenso de dados tem sido gerado que contrapõe a idéia de que a transcrição pervasiva de ncRNAs constitua um artefato. Além disso, foi observado em diversos trabalhos que a expressão de ncRNAs ocorre de uma maneira tecido-específica (Ravasi et al. 2006; Nakaya et al. 2007; Cabili et al. 2011), sugerindo que a transcrição desses RNAs seja regulada. Um trabalho interrogando 10 cromossomos humanos mostrou que 56% dos RNAs presentes no citoplasma não mapeavam em genes conhecidos, mRNAs ou ESTs (*Expressed Sequence Tags*) (Cheng et al. 2005), apontando para a existência de novos RNAs não codificadores e indicando que esses transcritos possam ser processados. Ainda, estudos da cromatina nas regiões genômicas que dão origem a esses ncRNAs revelaram assinaturas de histonas muito similares às regiões transcritas de genes codificadores de proteínas ativos como: trimetilação da lisina 4 da histona 3 (H3K4me3) na região promotora e trimetilação da lisina 36 da histona 3 (H3K36me3) no corpo do gene (Barski et al. 2007; Mikkelsen et al. 2007; Guttman et al. 2009). Além disso, sítios de ligação de fatores de transcrição (sp1, c-myc

e p53) já foram associados a transcritos não codificadores, e alguns desses responderam ao tratamento com acido retinóico, mostrando que a transcrição desses RNAs pode ser regulada (Cawley et al. 2004). Consistente com as evidências acima, o trabalho de Khalil e colaboradores (Khalil et al. 2009) identificou muitos lncRNAs ligados com *Polycomb* e outros complexos remodeladores da cromatina, sugerindo que esses RNAs possam de fato exercer papel importante em processos biológicos.

Tem sido convencional categorizar os ncRNAs em função do seu tamanho: RNAs longos (lncRNAs) possuindo tamanho maior do que 200 nucleotídeos e RNAs curtos, menores que 200 nucleotídeos (Clark and Mattick 2011), sendo este um critério arbitrário. Os ncRNAs curtos mais conhecidos são: miRNA (microRNA)[12], *small nuclear* RNAs (snRNAs)[13], *small nucleolar* RNAs (snoRNAs)[14], piwiRNA[15]. Outros ncRNAs curtos menos conhecidos são: RNAs iniciadores de transcrição (tiRNA), associados à região promotora (PASR) entre outros (Forrest et al. 2009). Os RNAs iniciadores de transcrição (tiRNAs) possuem em geral 18 nt e são transcritos pela RNA pol II. Estes RNAs podem ser subproduto da transcrição. Entretanto, acredita-se que tiRNAs sejam eventualmente formados quando a RNA pol II é forçada a recuar depois de se chocar com o primeiro nucleossomo do início de transcrição (Taft et al. 2009). RNAs associados aos promotores PARS (*promoter-associated short RNA*), são pequenos RNAs que parecem correlacionar com nível de expressão do codificador de

---

[12] miRNA: tipicamente gerados pela RNA pol II em unidades transcricionais independentes processados pela *Drosha* e *Dicer*, formando transcrito de aproximadamente 22nt para atuar como reguladores pós-transcricionais. Podem ser produzidos independente de *Drosha* a partir de introns pequenos, sendo produtos do processo de *splicing*.

[13] snRNAs: classe de RNAs curtos (~150 nt) encontrados dentro do núcleo de células eucarióticas, possuindo como principal função o processamento de pré-mRNAs no núcleo.

[14] snoRNAs: classe de pequenos RNAs tipicamente envolvidos no processamento de pré-rRNAs. A sua principal função se baseia na modificação química de nucleotídeos pela pseudouridilção (*H/ACA box snoRNAs*) ou metilação de bases específicas (*C/D box snoRNAs*).

[15] piwiRNAs(piwi-interacting RNAs): pequenos RNAs regulatórios (26 a 30 nt) conhecidos por interagirem com proteínas Ago da subfamília PIWI e atuam dirigindo o silenciamento de expressão de retrotransposons.

proteína. Por exemplo, o aumento da expressão de PASR relacionado ao *c-myc* diminuiu a expressão do respectivo codificador (Fejes-Toth et al. 2009).

A classe de lncRNAs inclui RNAs longos que possuem baixo ou nenhum potencial codificador e que podem mapear em regiões intra ou intergênicas. Os lncRNAs intragênicos podem ser representados por: transcritos antissenso, os quais são transcritos na fita oposta dos genes codificadores de proteína (Faghihi and Wahlestedt 2009; Wright and Bruford 2011), transcritos intrônicos, os quais possuem mesma orientação do codificador de proteína, mas se localizam apenas na região intrônica do *locus* (Louro et al. 2007; Wright and Bruford 2011), transcritos divergentes, os quais são transcritos de a partir do promotor do codificador de proteína na direção oposta (Rinn and Chang 2012). É possível encontrar isoformas alternativas de *splicing* que não codificam proteínas (Hashimoto et al. 2009; Chen et al. 2011) e também mRNAs que possuem papéis regulatórios que não traduzem para proteína (Candeias et al. 2008). Já os RNAs longos intergênicos (*lincRNAs*) são unidades transcricionais fisicamente separadas dos *loci* codificadores de proteína, definidas como localizando-se a no mínimo 5kb de um gene conhecido (Guttman et al. 2009; Wright and Bruford 2011; Rinn and Chang 2012).

Muitos ncRNAs apresentam estrutura parecida com os RNAs codificadores de proteínas como CAP 5’ e cauda poli(A)+, sendo denominados *mRNA-like* ncRNAs (Erdmann et al. 2001; Numata et al. 2003), a presença desses elementos pós-transcricionais é mais um indício de que esses ncRNAs não são ruído transcricional (Kavanaugh and Dietrich 2009). Entretanto, um trabalho recente mostrou que uma porção significante de ncRNAs do transcriptoma humano não é poliadenilada (Kapranov et al. 2010).

Diante de evidências de que o genoma é altamente transcrito e o DNA não funcional pode ser transcrito para gerar RNAs que parecem exercer funções celulares importantes

(Baker 2011), cientistas se juntaram para compor o projeto denominado ENCODE (*Encyclopedia of DNA Elements*) (Birney et al. 2007), o qual buscava caracterizar elementos funcionais na sequência de DNA do genoma humano. Ao término de uma etapa piloto do projeto que caracterizou 1% do genoma humano, estimou-se que pelo menos 90% do DNA humano seja transcrito. Ao témino do projeto ENCODE observou-se que esta taxa era de 75% (Consortium et al. 2012) e deu origem a uma iniciativa paralela, o GENCODE, focada na anotação de genes codificadores de proteínas e ncRNAs (Harrow et al. 2012).

1.3 LncRNAs e câncer

Alterações na expressão de lncRNAs tem sido documentada em vários tipos diferentes de cânceres humanos (Huarte and Rinn 2010; Gibb et al. 2011; Prensner and Chinnaiyan 2011). A expressão aberrante de *T-UCRs* (transcribed ultraconserved regions)[16] já foi identificada em leucemias e foi observado que alguns desses podem ser regulados por miRNAs. Além disso, foi observado que a expressão de *T-UCRs* encontra-se aumentada em neuroblastomas mais agressivos (Mestdagh et al. 2010). No estudo de Perez e colaboradores foi identificada a expressão aberrante de lncRNAs (>400 nt) em amostras de tumores de mama e ovário quando comparados ao tecido normal (Perez et al. 2008). Análises de 272 bibliotecas de SAGE identificou conjuntos de lncRNAs diferencialmente expressos (*fold change* > 2) em carcinomas de cérebro, mama e pulmão, respectivamente (Gibb et al. 2011).

[16] *T-UCRs* (*transcribed ultraconserved regions*): RNAs longos de no mínimo 200bp, transcritos a partir de regiões intragênicas que, aparentemente, não possuem potencial codificador, e precisam apresentar 100% de conservação entre humanos, camundongos e ratos (Bejerano et al. 2004; Calin et al. 2007).

Entre os poucos lncRNAs já bem caracterizados, *MALAT-1*[17] foi associado com metástase e mau prognóstico de pacientes com câncer de pulmão de células não pequenas (Ji et al. 2003). Ainda, este lncRNA pode interagir com PSF (*polypirimidine tract-binding protein-associated splicing factor*) para induzir a expressão do proto-oncogene *GAGE6* em humanos (melanoma) e camundongos (Li et al. 2009). Foi observado que o lncRNA *HOTAIR*[18] possui potencial preditor em relação à metástase e sobrevida em 5 anos em pacientes com câncer de mama; pacientes que apresentam maiores níveis de *HOTAIR* apresentaram pior prognóstico em relação à metástase e sobrevida (Gupta et al. 2010). Recentemente, *HOTAIR* também foi identificado como aumentado em amostras clínicas de pacientes com câncer de pâncreas em relação às amostras não tumorais, onde pacientes com expressão elevada deste transcrito apresentavam menor taxa de sobrevida em 5 anos (Kim et al. 2012). Este mesmo trabalho demonstrou que o silenciamento de *HOTAIR* em células tumorais de pâncreas (PANC-1 e L3.6pL) diminuiu a proliferação celular, alterou a progressão do ciclo celular e induziu a apoptose. Estudos independentes têm mostrado que a super-expressão do lncRNA *ANRIL*[19] em câncer de próstata promoveu o silenciamento de *INK4n/ARF/INK4a*, e *p15/CDKN2B* (Yu et al. 2008; Yap et al. 2010). O lincRNA *PCAT-1* foi identificado como aumentado em amostras tumorais de próstata (primário e metástase) comparada às amostras benignas (próstata e não próstata), o silenciamento deste transcrito

---

[17] *MALAT-1* (*Metastasis-associated Lung Adenocarcinoma Transcript-1- non-protein coding*) localizado no cromossomo 11q13 e possui 8.7 kb. Este lncRNA também pode ser clivado em um RNA menor denominado *masc* (*MALAT1-associated small cyotplasmatic RNA*) que é exportado para o citoplasma (Wilusz et al. 2008). Este transcrito também já foi descrito como importante para manter a estrutura dos *speckles* e regular fatores de *splicing* (Tripathi et al. 2010).

[18] *HOTAIR*: RNA não codificador longo de 2.2kb que é expresso a partir do *locus HOXC* atua em genes do *locus HOXD* por recrutar o complexo PCR2 para inserir as marcas de cromatina H3K27me3 reprimindo a expressão (Rinn et al. 2007).

[19] *ANRIL*: transcrito antissenso de tamanho 2.2kb. Este transcrito controla a expressão do *locus* gênico *INK4/ARF*, composto por genes supressores de tumor (*INK4n/ARF/INK4a*, *p16/CDKN2A* e *p15/CDKN2B*) por meio da interação com CXB7 (componente do complexo repressivo *polycomb* 1, PRC1) estabelecendo marcas epigenéticas de repressão (H3K27me3) (Yap et al. 2010).

diminui a proliferação celular e parece ser regulado por EZH2 (proteína do complexo *polycomb* PCR2) (Prensner et al. 2011).

Alguns lncRNAs tem sido investigados como marcadores de diagnóstico de câncer. Um exemplo é *PCA3*, que pode ser identificado na urina de pacientes de câncer de próstata, e apresenta uma maior especificidade e sensibilidade do que o marcadores mais usuais como o PSA (*prostate-specific antigen*) (Tinzl et al. 2004). O lncRNA *HULC* é altamente expresso em pacientes com hepatocarcinoma e pode ser detectado no sangue por métodos de PCR convencional (Panzitt et al. 2007). Como mostrado na seção 1.1, os miRNAs já foram identificados com expressão alterado em PDAC, e se apresentam promissores para o seu uso no diagnóstico da doença (Hamada and Shimosegawa 2011; Steele et al. 2011). Entretanto, ainda falta informação na literatura abordando a expressão de lncRNAs no adenocarcinoma pancreático. Apenas, recentemente, o lncRNA HOTAIR foi correlacionado a esse tipo de câncer (Kim et al. 2012), e no trabalho realizado por Ting e colaboradores, foi observada a superexpressão de RNAs satélites (HSATII) de tamanho de 100 a 500 nt em pacientes com PDAC (Ting et al. 2011). Interessantemente, a detecção de HSATII por hibridação *in situ* de RNA foi capaz de confirmar o diagnóstico de PDAC em biopsias tumorais, incluindo casos em que a histopatologia não tinha sido identificada (Ting et al. 2011). Desta forma, fica claro que os lncRNAs são transcritos funcionais e podem desempenhar papel importante no processo de tumorigênese e possuem potencial para serem utilizados como marcadores moleculares para a detecção de tumores humanos.

2 Objetivos

2.1 Objetivos gerais

Investigar a expressão de RNAs longos não codificadores originados de regiões intrônicas e intergênicas do genoma em amostras de tecido pancreático humano com diferentes graus de malignidade

2.2 Objetivos específicos

- Gerar perfis de expressão gênica a partir de amostras clínicas de adenocarcinoma de pâncreas, metástases de tumores pancreáticos, pancreatite crônica e tecido pancreático não-tumoral utilizando microarranjos de cDNA contendo sondas para lncRNAs e mRNAs codificadores de proteínas;
- Identificar lncRNAs e mRNAs com expressão diferencial em adenocarcinomas primários e metástases de tumores de pâncreas;
- Identificar vias moleculares ou processos biológicos com expressão alterada no adenocarcinoma pancreático primário e em metástases;
- Caracterizar *in silico* propriedades globais dos lncRNAs detectados em tecidos pancreáticos: expressão em outros tecidos, tamanho do transcrito, potencial codificador e potencial precursor de pequenos RNAs;
- Investigar a presença na sequencia genômica de possíveis regiões regulatórias de lncRNAs expressos em tecidos pancreáticos;
- Validar e caracterizar lncRNAs que compõem a assinatura de expressão associada à metástase pancreática utlizando qPCR.

3 Materiais e Métodos

3.1 Coleta e armazenamento de amostras

As amostras tumorais de tecidos pancreáticos foram coletadas de pacientes diagnosticados com câncer de pâncreas e submetidos à ressecção cirúrgica do tumor no Hospital das Clínicas, Faculdade de Medicina – USP (FMUSP) pelo grupo do Prof. Dr. Marcel C. C. Machado. Inicialmente, todas as amostras utilizadas neste estudo foram obtidas com consentimento dos pacientes e aprovação do Comitê de Ética do Hospital das Clínicas – FMUSP para a utilização neste estudo.

As amostras foram congeladas entre 5 e 10 min após a ressecção, sendo conservadas em nitrogênio líquido no banco de tecidos frescos. As amostras tumorais foram submetidas à macro-dissecação, ainda mantidas congeladas, para remover o tecido não tumoral adjacente. O projeto inicial era composto por 39 amostras tumorais, incluindo tecidos pancreáticos de tumores primários e tecidos metastáticos (biópsias de outros órgãos), dezessete amostras de tecidos adjacentes ao tumor e 12 pancreatites crônicas. Os pacientes que possuíam pancreatite crônica foram operados no mesmo hospital, pelo mesmo grupo, seguindo as recomendações apresentadas anteriormente. Quarenta amostras resultaram em RNA de boa qualidade e foram utilizadas nos experimentos de expressão gênica. Os dados clínicos e anatomopatológicos das amostras analisadas são apresentados na Tabela 1. Posteriormente aos experimentos, duas amostras (215T e 10N) foram re-anotadas como sendo adenocarcinomas de via biliar e tecido não-neoplásico adjacente, sendo então removidas das análises.

Tabela 1: Dados clínicos dos 38 pacientes que tiveram suas amostras analisadas

N° do Paciente	Sexo	Idade	Óbito	Tipo Histológico	Sítio metastático
1	F	51	2	Normal	-
2	F	51	0	Normal	-
3	ND	ND	ND	Normal	-
4	M	63	0	Normal	-
5	F	72	2	Normal	-
6	F	ND	ND	Normal	-
7	M	ND	ND	Normal	-
8	M	ND	ND	Normal	-
9	ND	ND	ND	Normal	-
101	M	35	ND	Pâncreatite Crônica	-
102	M	50	ND	Pâncreatite Crônica	-
103	M	62	ND	Pâncreatite Crônica	-
104	M	ND	ND	Pâncreatite Crônica	-
105	ND	ND	ND	Pâncreatite Crônica	-
106	M	53	0	Pâncreatite Crônica	-
107	M	45	0	Pâncreatite Crônica	-
108	M	48	ND	Pâncreatite Crônica	-
201	F	51	2	Adenocarcinoma pâncreas primário	-
202	ND	ND	ND	Adenocarcinoma pâncreas primário	-
203	M	57	0	Adenocarcinoma pâncreas primário	-
204	F	65	2	Adenocarcinoma pâncreas primário	-
205	M	46	ND	Adenocarcinoma pâncreas primário	-
206	ND	ND	ND	Adenocarcinoma pâncreas primário	-
207	F	60	0	Adenocarcinoma pâncreas primário	-
208	F	65	0	Adenocarcinoma pâncreas primário	-
209	ND	ND	ND	Adenocarcinoma pâncreas primário	-
210	ND	ND	ND	Adenocarcinoma pâncreas primário	-
211	ND	ND	ND	Adenocarcinoma pâncreas primário	-
212	F	51	0	Adenocarcinoma pâncreas primário	-
214	M	ND	ND	Metástase	Fígado
216	F	72	2	Adenocarcinoma pâncreas primário	-
217	F	49	0	Adenocarcinoma pâncreas primário	-
218	M	ND	ND	Adenocarcinoma pâncreas primário	-
302	F	76	1	Metástase	Peritôneo
303	F	83	0	Metástase	Fígado
304	F	55	2	Metástase	Fígado
305	ND	ND	ND	Metástase	Fígado
306	ND	ND	ND	Metástase	Gânglio

ND: não definido, F: feminino, M: masculino, 0: não óbito, 1: óbito não realcionado à doença, 2: óbito relacionado à doença.

3.2 Extração de RNA das amostras clínicas de pacientes

Para extração de RNA total, as amostras de tecido ainda congeladas em nitrogênio líquido foram pulverizadas em um almofariz com ajuda de um pistilo e a seguir processadas pelo método de uso do tiocianato de guanidina fenol-clorofórmio. A extração das amostras foi realizada por TRIzol®(Invitrogen), seguido pela purificação do RNA total obtido em colunas RNeasy Mini *Kit* (Qiagen). O TRIzol é um reagente monofásico composto por fenol e isotiocianato de guanidina. Durante a homogeneização e a lise celular, o reagente mantém a integridade do RNA. As amostras foram maceradas em N_2 líquido e transferidas para tubos com TRIzol. Para cada 50mg de tecido foi utilizado 1 ml de TRIzol. Após 5 min de incubação à temperatura ambiente para permitir a completa dissociação dos complexos ribonucleoprotéicos, 200µl de clorofórmio foram adicionados para cada 1ml de TRIzol inicial. Após uma homogeneização vigorosa, as amostras foram centrifugadas a 11000g por 15min a 2°C, para que a separação da solução ocorresse em 2 fases: aquosa (incolor e superior) e orgânica (rosa e inferior). O RNA foi precipitado a partir da fase aquosa, misturando 500µl de isopropanol para cada 1ml inicial de TRIzol. As amostras foram incubadas por 10min a temperatura ambiente e centrifugadas a 11000g a 2°C por 10 min. O sobrenadante foi removido e o *pellet* de RNA formado no fundo do tubo foi lavado com etanol 75% (7000g a 2°C por 5 min). O sobrenadante foi removido e o pellet foi seco a temperatura ambiente e eluído em água tratada com DEPC (dietil pirocarbonato). A quantidade de RNA obtida foi medida utilizando o espectrofotômetro ND-100 (Nanodrop). A integridade do RNA foi avaliada por eletroforese microfluídica utilizando o equipamento Bioanalyzer 2100 (Agilent). Este equipamento realiza uma corrida de eletroforese em escala microfluídica com capacidade para 12 amostras em cada corrida no chip (Nano). Nessa técnica, a presença ou ausência das subunidades 28S e 18S dos RNAs ribossomais são determinadas através da intensidade de fluorescência nos picos correspondentes à ~4000pb e

~2000pb, respectivamente. Idealmente, as bandas 28S e 18S, correspondentes às subunidades ribossomais maior e menor, respectivamente, deveriam estar presentes em uma proporção de aproximadamente 2:1 em uma amostra de RNA sem qualquer degradação. A integridade é avaliada pela taxa de detecção entre o RNA 28S e 18S que deve ser próxima de 2, entretanto este valor pode variar dependendo da amostra utilizada. Os RNAs ribossômicos são os mais abundantes dentro da porção de RNA total, logo, parte-se do princípio se apresentam íntegros, as outras populações também estarão. A avaliação da qualidade de amostra de RNA também foi realizada pelo Nanodrop ND 1000 por espectrofotometria, seguindo a lei de Beer-Lambert, para a quantificação da medida óptica a 260nm (ácido nucléicos), 230nm (etanol e etc.) e 280 nm (proteína). Para amostras de RNA com qualidade as razões da quantificação 260/230 e 260/280 precisam estar próximas de 2. Entretanto, esta taxa pode variar de acordo com a amostra.

Após a extração, as amostras de RNA foram submetidas ao tratamento com DNAse para eliminação de possível DNA genômico contaminante e purificadas utilizando o *kit* RNeasy Mini (QIAGEN). Esta etapa é muito importante, uma vez que o objeto de estudo eram lncRNAs de regiões intrônicas e intergênicas, assim a presença de DNA genômico poderia provocar o aparecimento de um sinal falso positivo nas análises. O procedimento utilizado pelo *kit* representa uma tecnologia bem estabelecida para a purificação de RNA, combinando propriedades de ligação seletiva de RNA à coluna de sílica e a velocidade de centrifugação. Para realizar a etapa de purificação, as amostras tiveram o seu volume completado para 100µL. Na primeira etapa, foram adicionados 350µL de tampão RLT, que contem uma solução salina de isotiocianato de guanidina que possui propriedades desnaturantes, em alguns casos β-mercaptoetanol foi adicionado na proporção 1µL para cada 100µL de tampão RLT para aumentar a propriedade desnaturante, e ainda 250µL de etanol 100% para promover condições de ligação entre o RNA e a sílica. Os 700µL foram

adicionados à coluna centrifugados durante 15s a 8000g. As colunas foram lavadas com 350µL de tampão RW1, que continha etanol e pouca quantidade de isotiocianato de guanidina, durante 15s a 8000g. Logo em seguida foram adicionados na coluna solução preparada com 10µL de DNAse I (QIAGEN) e 70µL de RDD mantidas a temperatura ambiente por 15min, o RNA era tratado com DNAse I enquanto estava ligado à coluna de sílica, Qiagen. Novamente, as colunas foram lavadas com 350µL de RWI e centrifugadas a 8000g por 15s, seguidas de duas lavagens com tampão RPE a 8000g por 15s na primeira lavagem e 2min na segunda lavagem. Após estas lavagens, as colunas foram centrifugadas a 16000g por 1min para eliminar a maior parte dos resíduos de etanol. Para eluir o RNA da coluna, foram adicionados 30-40µL de água tratada com DEPC para hidratar a coluna de sílica e centrifugadas a 8000g por 1 min. Em alguns casos esse último passo foi realizado novamente.

A integridade do RNA foi avaliada por eletroforese microfluídica utilizando o Bioanalyzer 2100 (Agilent). Do total de amostras disponíveis, 22 amostras (~32% do total) apresentaram RNA degradado, e não foram utilizadas nas análises subsequentes. Tendo em vista que o pâncreas constitui um tecido extremamente enriquecido em ribonucleases (Reddi and Holland 1976), a obtenção de RNA de boa qualidade de mais de 67% das amostras clínicas disponíveis pode ser considerada um índice de aproveitamento satisfatório. Como exemplo, na Figura 2 está apresentado o resultado da separação por eletroforese capilar de 10 amostras de RNA isolado de tecido pancreático.

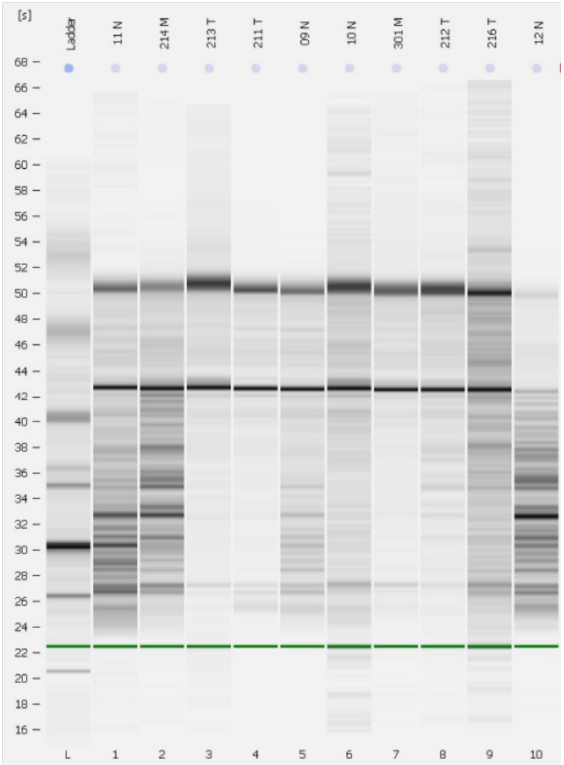

Figura 2: Análise por eletroforese capilar de alíquotas de RNA total de pâncreas isolado de amostras clínicas. Canaletas de 1 a 9: notar as bandas 28S (~4000 nt) e 18S (~2000 nt) de amostras com qualidade aceitável (i.e. razão 28S/18S ≥ 1) para análise subsequente. Canaleta 10: amostra com degradação intensa e excluída do estudo. Esta figura representa uma imagem de um gel virtualmente construído pelo *software*, as bandas são referentes à área dos picos registrados pelo detector.

A maior parte das amostras apresentou parâmetros de RIN (*RNA Integrity*) e relação [28S/18S] bons. O parâmetro RIN é definido pelo próprio *software* do Bioanalyzer que leva em consideração a presença dos picos referentes às unidades ribossomais (18S e 28S) e também o padrão de degração dos RNAs, podendo variar de 0 a 10. No caso de amostras clínicas, a literatura descreve uma dificuldade em obter razões 28S/18S próximas a 2, devido a manipulação do material antes do armazenamento da amostra (congelamento em nitrogênio) (Blackhall et al. 2004; Ohashi et al. 2004). Para as amostras que obtiveram a relação [28S/18S]<1 para amostras clínicas de tecido pancreático foram descartadas.

3.3 Avaliação da contaminação genômica (DNA) no RNA total

Para verificar a ausência de DNA genômico contaminante na amostra de RNA total após o tratamento com DNAse I, foram realizadas reações de *PCR* (*Polymerase Chain Reaction*). As amostras foram descongeladas e colocadas em gelo, para inibir a degradação do RNA. Em cada reação foi adicionada um *mix* contendo 5µl de tampão (10x), 1,3µl de MgCl (10mM), 1µL de dNTP (10mM), 1µL de *Primer Forward* e 1µL de *Primer Reverse* para o gene da Tubulina (10mM), expresso constitutivamente em células do organismo, 0,5µL de enzima Taq Biolase (Biolase) e 39,2µl de Água DEPC, totalizando uma reação de 50µL.

Foram utilizados controles positivo e negativo na reação de PCR. O controle negativo (reação sem amostra) foi feito para verificar a ausência de contaminação com DNA nos reagentes, garantindo que a detecção da presença de amplificação do DNA foi proveniente apenas da amostra, evitando um resultado falso-positivo; e o controle positivo (reação utilizando DNA genômico como molde) foi feito para verificar a integridade e eficiência de todos os reagentes do *kit*. A reação foi realizada em 40 ciclos, que eram constituídos por 3 etapas, conforme recomendação do fabricante:

1ª etapa: desnaturação da fita de DNA, ocorreu a 95°C por 30s, para desnaturar a fita de DNA e evitar a formação de estruturas secundárias;

2ª etapa: anelamento dos iniciadores, em 55°C por 30s;

3ª etapa: alongamento da cadeia, em 72°C por 30s, temperatura ótima para a reação de polimerização realizada pela Taq Polimerase, a vantagem de utilizar alta temperatura para elongação da cadeia possui maior probabilidade de evitar a formação de estruturas secundárias que impediriam a formação completa do transcrito.

O resultado obtido na PCR foi visualizado por eletroforese em gel de agarose. Para isso foi preparado gel de agarose 2%, a porcentagem escolhida está relacionada com a dimensão do fragmento a ser visualizado, neste caso 400pb, correspondente ao gene da tubulina. Para cada 400 mg de agarose foram adicionados 20 ml de tampão TAE (Tris, ácido acétido e EDTA) e 1µl de brometo de etídio. O ladder (padrão) utilizado foi *Gene Ruler* 50bp (SM0371,Fermentas). Para visualização da banda, foi adicionado brometo de etídio ao gel de agarose numa proporção de 1,0µl para cada 20ml de gel de agarose. A Figura 3 representa o resultado típico da eletroforese do PCR utilizado para verificar a possível contaminação com DNA genômico. Os resultados são representativos das amostras utilizadas no projeto.

Figura 3: Verificação de ausência de contaminação por DNA genômico nas amostras de RNA que foram submetidas à PCR com *primers* de tubulina (200pb). Na primeira coluna (à esquerda) foi aplicado o padrão de 100pb, as colunas subseqüentes correspondem às amostras 07N, 206T, 205T, 202T, 003N, 101PC, 203T, 208T, 209T, 102PC, 04N, 104PC, 05N, 06N. Na 15°coluna foi aplicado o controle positivo, o DNA genômico pode ser visualizado pela banda em aproximadamente 200pb. A última coluna corresponde ao controle negativo (sem amostra).

3.4 Amplificação linear no RNA mensageiro (mRNA)

O processo de amplificação linear de mRNA foi realizado para que fosse possível incluir no estudo amostras que resultaram na obtenção em quantidade insuficiente de RNA para a marcação direta de alvos fluorescentes. Para contornar esta limitação, foi utilizado o protocolo de Eberwine, (Van Gelder et al. 1990; Eberwine 1996). Já foi demonstrado que este protocolo garante a manutenção da representação do transcriptoma no RNA amplificado (Wang et al. 2000), e já se encontrava padronizado em nosso laboratório. Este método envolve dois ciclos

de amplificação do mRNA e envolve a formação de cDNA pela transcrição reversa pela enzima transcriptase reversa (Superscript III, Invitrogen) utilizando *primer* específico oligo dT-T7 acoplado com o sítio para a T7 RNA polimerase (T7 Megascript *Kit* – Ambion). A síntese da primeira fita de cDNA foi realizada utilizando *primers* oligo dT-T7 para a seleção de mRNAs partindo de 1µg de RNA total.

Na primeira etapa, 1µg de RNA total, tratado com DNAse, foi aliquotado em 9µL de água DEPC, 1µL de dT T7 oligo *primer* (0,5µg /µL) foi adicionado e a mistura incubada a 65ºC por 5 minutos para desfazer as estruturas secundárias do RNA. Em seguida foi adicionado um *mix* contendo 4µL de tampão 5X *First Strand*, 2µL de dNTP (10mM), 1µL da enzima RNasin ribonuclease inhibitor para inibir a atividade das RNases e evitar que o RNA fosse degradado e 2µL da enzima Superscript III. A mistura foi incubada a 46ºC por 90 minutos para a síntese da primeira fita de cDNA. Após 20min de transcrição, foi adicionado 1µL de *TS Primer* (0,5µg/µL), denominado *template switcher* (5'-AAG CAG TGG TAT CAA CGC AGA GTA CGC GGG-3') que se liga na porção CCC localizada na ponta 3’ do cDNA. Para a síntese da segunda fita de cDNA, um *mix* contendo 105µL de água DEPC, 15µL de Tampão *Advantage* PCR, 3µL de TS *Primer* (0,5µg/µL), 3µL de dNTP (10mM) e 3µL das enzimas *Advantage cDNA Polymerase mix* e RNase H (2U/µL) foram adicionados à reação. A enzima RNase H degrada o RNA molde e a DNA polimerase presente no *mix* é responsável pela extensão da cadeia da 2ª fita de cDNA. A mistura foi colocada no termociclador programado com as temperaturas a 37ºC por 5 min para a digestão do mRNA molde, em seguida a temperatura foi elevada para 94ºC, durante 2 min para a desnaturação, na etapa seguinte, a temperatura foi abaixada para 65ºC e mantida nesse valor por 1 min para o anelamento específico dos *primers* e logo depois foi elevada para 75ºC temperatura ótima de atividade da enzima.

O cDNA dupla fita formado foi purificado (mostrado no item 3.4.1) e a transcrição foi realizada pela adição de 2µl de cada NTP (A, U, G e C), 2µl de tampão e 2µl de um *mix* contendo a enzima T7 RNA polimerase e inibidor de RNase (*kit* Ambion). A amplificação do cRNA foi realizada a 37ºC durante 6 horas. Após 6 horas de transcrição *in vitro*, cada solução foi transferida para um tubo livre de RNase contendo 1 mL de TRIzol, reagente composto por solução monofásica de isotiocianato de guanidina e fenol que mantém a integridade do RNA. Foram adicionados 200µL de clorofórmio em cada tubo e as amostrar foram mantidas por 2-3 minutos em temperatura ambiente. Os tubos foram centrifugados a 4ºC com rotação de 13000rpm por 15 minutos. A adição de clorofórmio seguida de centrifugação permitiu a separação da solução em fase aquosa e fase orgânica, o RNA permaneceu na fase aquosa e a maior parte do DNA e das proteínas fica retida na fase orgânica. Com o término da centrifugação, a fase aquosa, foi transferida para outro tubo de livre de RNase contendo 500µL de isopropanol e a fase orgânica descartada. A solução foi homogeneizada e mantida a temperatura ambiente durante 10 min. Em seguida os tubos foram centrifugados a 4ºC por 25 minutos com rotação de 13000rpm com o propósito de precipitar totalmente o cRNA total produzido. A formação de *pellet* evidencia a presença de cRNA. O sobrenadante de cada tubo foi removido, 800µL de etanol 70% foram adicionados, e foi realizada centrifugação a 4ºC durante 10 min com rotação de 13000rpm. Após a centrifugação, o sobrenadante foi removido e a lavagem repetida. Após evaporação total do etanol, o cRNA foi ressuspendido em 20µL de água DEPC. Nesse ponto a quantidade e qualidade do cRNA I obtido no primeiro ciclo de amplificação foi verificada pela medida da razão de absorbância a DO260/280 (~2) e por eletroforese microfluídica (Bioanalyzer).

Para realizar o segundo ciclo de amplificação, o cRNA obtido foi purificado. A síntese da primeira fita foi realizada da mesma forma que no primeiro ciclo, utilizando 1µg do cRNA obtido no primeiro ciclo de amplificação, com exceção que, o dT-T7 oligo *primer* foi

substituído por dN6 (random hexamer), pois a seleção dos mRNAs ocorreu na primeira etapa da amplificação. Na síntese da segunda fita de cDNA, 3μL de TS *primer* foram substituídos por 4μL de dT-T7 oligo *primer* (0,5μg/μL) e as temperaturas de anelamento e extensão da cadeia foram alteradas para 45ºC e 68ºC, respectivamente. Essa alteração foi feita pela diminuição da estringência de anelamento de oligo dT-T7, pois todo o mRNA já havia sido selecionado no primeiro passo de amplificação. O cDNA dupla fita foi purificado segundo o item 3.4.1. Na segunda etapa da transcrição, o volume dos NTPs (A, G e C), do *mix* de enzima e do tampão de reação foram dobrados, o volume de UTP permaneceu o mesmo, porque nessa etapa, foram adicionados 2μL de aminoalil UTPs (aaUTP).

Após a reação de transcrição *in vitro* durante 13 horas, as amostras foram tratadas com DNAse e a purificação realizada com o RNeasy Mini *Kit* (Qiagen). A coluna do *kit* contém uma membrana de sílica capaz de reter o RNA na presença de isotiocianato de guanidina, sendo extremamente eficiente em processos de purificação de amostras. As amostras tiveram o volume ajustado para 100μL e a massa de RNA esperado não poderia exceder 100μg, capacidade máxima de cada coluna de purificação. Todo o procedimento foi realizado a temperatura ambiente, conforme recomendação do fabricante.

Em cada amostra, foi adicionado um *mix* que continha 350μL de solução tampão RTL, contendo isotiocianato de guanidina, e 250μL de etanol (96-100%) totalizando 700μL de solução formada que foram inseridos dentro da coluna Rneasy. O *mix* preparado tem a função de criar um ambiente propício para a formação da ligação específica entre o RNA e a membrana de sílica da coluna. O tubo com a solução (*mix* + amostra) foi centrifugado durante 15s a 8000g. Após esta etapa, o filtrado foi descartado e 350μL de tampão RW1 foram inseridos na coluna, e, novamente o tubo foi centrifugado 15s a 8000g. Após a centrifugação, 500μL de tampão RPE, também um tampão de lavagem, foram inseridos na coluna duas

vezes. Na primeira, houve a centrifugação durante 15s a 8000g e na segunda a centrifugação foi realizada durante 2min a 8000g. Depois de ter realizado as etapas acima, a coluna foi centrifugada durante 1 minuto na velocidade máxima da centrífuga para que o etanol evaporasse completamente, pois qualquer resíduo deste álcool poderia atrapalhar na dissolução do RNA. Para a eluição, foram adicionados na coluna, 40µL de água livre de RNase. O sistema foi centrifugado durante 1min a 8000g. Consideramos adequadas para hibridização as amostras amplificadas em que se observou um rastro (*smear*) com tamanho na faixa de 2.000 a 200 nucleotídeos no Bioanalyzer. A quantidade mínima requerida de cRNA de cada amostra é de 10µg, necessários para realizar 2 hibridizações (réplicas técnicas), utilizando 5µg em cada uma delas.

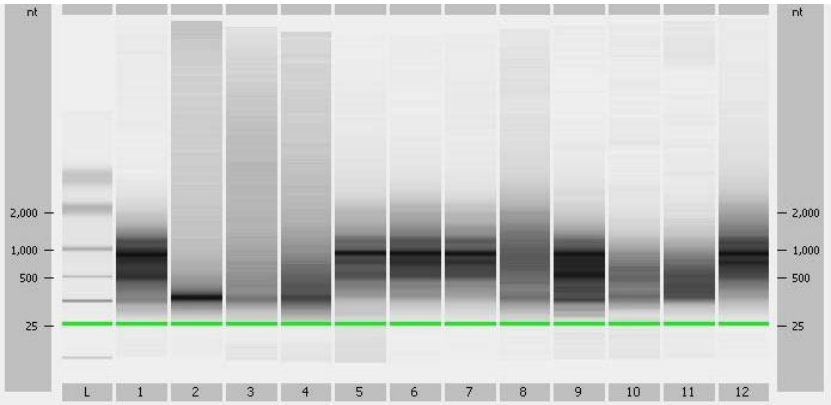

Figura 4 Exemplo típico de resultado de corrida de cRNA em gel no 2100Bioanalyzer. Apenas amostras onde se observa um *smear* predominante entre 200−2.000 nt foram utilizadas nas hibridizações. Note que as amostras nas canaletas 2, 3 e4 mostram uma banda de amplificação na altura dos 200 nt, entretanto, não apresentam RNAs muito longos, logo tiveram que ser amplificadas novamente.

O cDNA dupla fita formado na primeira e segunda etapas da amplificação foi purificado antes da síntese de cRNA para evitar que dNTPs não incorporados, rNTPs resultantes da degradação do RNA, *primers* e enzimas inativadas interferissem na transcrição *in vitro*.

Para realizar a purificação foram utilizados tubos Phase Lock Gel (Eppendorf) para reter a enzima RNAse H no gel. Os tubos foram centrifugados a 13000rpm por 1min para que o gel se depositasse no fundo e 160µL de uma solução fenol-clorofórmio-isoamil foram adicionados. A cada um dos tubos contendo as amostras foram adicionados 16µL de acetato de sódio 3M (pH 5,2) e a mistura foi transferida para o tubo Phase Lock Gel. A nova mistura foi homogeneizada e centrifugada por 6 min a 13000rpm, foi observada a formação de três fases, a orgânica, inferior, a fase em gel, na qual estava aderida a enzima e a fase aquosa, superior, que continha o material de interesse. A fase superior foi transferida para um tubo RNAse free e foram adicionados 1,2µL de glicogênio (20mg/ml) e 320µL de etanol 100%. Os tubos foram incubados por 20min a -20ºC e em seguida centrifugados a 14000rpm por 20min. A formação de um pequeno *pellet* branco indica a presença de cDNA dupla fita. O *pellet* foi lavado com 500µL de etanol 100% e centrifugado por 5min a 14000rpm. O sobrenadante foi descartado e esperou-se que todo o etanol restante fosse evaporado. O cDNA dupla fita foi ressuspendido em 20µL de água DEPC.

Para a obtenção de cDNA bastante puro, foram utilizadas colunas de purificação NUCAWAY (Ambion), composta por polímero que adsorve oligonucleotídeos. Para a preparação da coluna, foi adicionado 1mL de água DEPC e realizada homogeneização para que fossem eliminadas bolhas. As colunas foram centrifugadas a 735g por 2min para retirar o excesso de água. As amostras foram aplicadas nas colunas e estas foram centrifugadas a 735g por 2min. O eluído obtido foi secado utilizando Speedvac para atingir o volume de 8µL.

3.4.2 Marcação de alvos fluorescentes a partir de cRNA com Cy5

Para a marcação de alvos fluorescentes representando o transcriptoma das amostras de pâncreas, foram utilizados 5µg de cRNA obtido na segunda etapa de amplificação. O RNA foi

concentrado por *Speedvac* e ressuspendido em 4.5μL de tampão de acoplamento (0.1M tampão carbonato pH 8.5-9.0). Foram adicionados 4μL de corante monoreativo de cianina 5 (Cy5) e a reação ocorreu por 1h a temperatura ambiente. Todo o processo foi realizado no escuro, para evitar a degradação do fluoróforo. Em seguida, 6μL de Hidroxilamina 4M foram adicionados para finalizar a reação. Isso ocorre porque a hidroxilamina reage com o corante que não acoplou com o cRNA, na fase anterior.

As amostras acopladas foram purificadas utilizando o *kit* da Qiagen *RNEasy clean-up*, para retirar o corante que não foi utilizado na reação de acoplamento, e, a densidade óptica das amostras foi medida pelo Nanodrop a 260/550/650 nm. A seguir, foram concentradas para um volume de aproximadamente 10μl utilizando colunas YM-30 Millipore. As amostras que não conseguiram atingir um volume de aproximadamente 10μl foram concentradas adicionalmente utilizando *SpeedVac*.

3.5 Hibridização de cRNA com microarranjos de DNA

O microarranjo de cDNA utilizado nos experimentos de hibridização foi construído em nosso próprio laboratório em 2002, no começo do projeto CAGE-Câncer . Este microarranjo contém 3355 fragmentos de cDNA que foram selecionados através de uma análise informática das 1.016.501 ESTs provenientes de bibliotecas de cDNA gerados a partir de mRNAs poliA+ de 20 tipos de tumores humanos, geradas a partir da técnica ORESTES entre 2000-2001 durante o Projeto Genoma Humano do Câncer (HCGP - http://www.ludwig.org.br/ORESTES) (Dias Neto et al. 2000; Camargo et al. 2001). O microarranjo contém também 30 controles positivos (*house keeping genes*) e negativos (regiões intergênicas de levedura e genes de bactéria e planta sem similaridade com transcritos humanos). Transcritos selecionados para compor a plataforma de microarranjos

foram anotados como codificadores de proteína, lncRNAs de regiões intrônicas ou intergênicas de acordo com o mapeamento no genoma humano (montagem hg19) e de acordo com a base de dados de RefSeq mRNA (http://genome.ucsc.edu/). Os lncRNAs provenientes de regiões intrônicas e intergênicas interrogados pela plataforma de microarranjos foram randomicamente escolhidos da base de dados e ESTs. Transcritos anotados como lncRNAs intrônicos (n = 722, i.e. 21%) compõem sequências que mapearam dentro de uma região intrônica de um codificador de proteína. Transcritos anotados como intergênicos (n = 188, i.e. 6%) representam sequências que mapearam no genoma, entretanto não possuem anotação de gene pela base de dados do RefSeq. Para serem considerados como intrônicos ou intergênicos a EST predita não podia mapear em nenhuma região exônica de genes conhecidos. Os transcritos anotados como lncRNAs conhecidos (n =74, i.e. 2%) se referem aos RNAs que possuem coordenadas genômicas que se sobrepõem às coordenadas de ncRNAs que estão anotados na base de RefSeq (anotados como NR). Transcritos anotados como mRNAs codificadores de proteína se sobrepunham à regiões exônicas de genes codificadores de proteínas. Para considerar a possibilidade de retenção de intrônica ou novos exons não anotados, transcritos que mapeavam parcialmente em regiões intrônicas e exônicas foram anotados como exônicos.

Para cada amostra analisada, uma alíquota de alvos fluorescentes foi ressuspendida em 200µl de solução tampão de hibridização (1x tampão de hibridização GE Healthcare V2, 50% Formamida). O tampão foi adicionado aos alvos, que sofreram o processo de desnaturação por 2 minutos à 92ºC sendo em seguida colocados no gelo para manter a desnaturação. Os tubos foram centrifugados durante 3 minutos à 8000g para que qualquer contaminante permanecesse no fundo do tubo.

Com uma seringa, cada amostra foi colocada em uma câmara de hibridização do hibridizador automático de lâminas (ASP – GE Healthcare). A reação de hibridização ocorreu durante 15 horas. Passado este tempo, as plataformas de miroarranjos (lâminas) foram lavadas com os tampões composto por SSC e SDS nas seguintes concentrações; 1x SSC/0,2% SDS, 0,1x SSC/0,2 %SDS e o último continha apenas 0,1x SSC. As lâminas foram secas com isopropanol. Em seguida, as lâminas hibridizadas foram analisadas em *scanner* confocal (Generation III Confocal Scanner − Molecular Dynamics) com a voltagem do PMT (PhotoMultiplier Tube) em cada fonte de laser ajustada a 600 V. As imagens obtidas foram processadas para extração dos dados de intensidade com o programa *ArrayVision 8.0* (Imaging Research, Canadá), resultando em planilhas em formato texto contendo os dados brutos de intensidade e os valores de intensidade subtraídos do *background* local.

3.6 Análise informática dos dados de expressão gênica

Os dados de intensidade utilizados foram aqueles que representavam os valores de intensidade bruta subtraído do respectivo *background*. Para trabalhar com as intensidades mais robustas medidas nos microarranjos, para cada réplica foram filtrados e descartados da análise *spots* cujos valores se encontravam abaixo da mediana (50% de valores menos intensos em cada lâmina), calculada entre os 4608 *spots* existentes. Para tornar os experimentos comparáveis, os dados referentes às diferentes réplicas de amostras foram normalizados pelo método do quantil (Bolstad et al. 2003). Na normalização por quantil, os *spots*, em cada lâmina, são ordenados (*ranked*) segundo o valor de intensidade. Estes valores são então substituídos pela média de intensidade obtida para determinada posição no *ranking*, obtendo-se assim um conjunto único de dados para todas as lâminas (Bolstad et al. 2003). Por esse método, é mais relevante a posição do *spot* no *ranking* do que a sua intensidade bruta. Com esse critério, pode-se comparar diferentes experimentos com escalas

de intensidades diferentes. Após esta etapa, a média de cada transcrito entre as réplicas foi calculada. Foram consideradas válidas as medidas de intensidade com valor superior a média do background mais 1 desvio-padrão.Utilizando o nº de réplicas que cada amostra possuía como critério de *cut-off*, amostras que possuíam 2 ou 3 réplicas deveriam apresentar no mínimo 2 réplicas válidas e aquelas que possuíam 4 réplicas, deveriam apresentar no mínimo 3 réplicas válidas. Do total de 4.608 de sondas representando transcritos humanos depositadas na lâmina, 1.607 (35%) foram consideradas válidas para análises subsequentes, sendo detectadas em pelo menos 75% das amostras em ao menos um dos tipos histológicos de pâncreas analisados.

Após o processo de filtragem e normalização, os dados foram processados utilizando o *script* ComBat.R (Johnson et al. 2007), disponível em http://statistics.byu.edu/johnson/ComBat/ dentro do ambiente R (http://www.r-project.org/). Esta etapa foi necessária pelo fato das amostras terem sido hibridizadas em dois lotes de lâminas fabricadas em períodos diferentes. Assim, os perfis de expressão gênica das plataformas de microarranjos agrupavam-se, predominantemente, em função do lote de hibridização, causando um viés na análise que impossibilitava uma análise conjunta dos dados. Efeitos associados aos diferentes lotes de hibridização têm sido observados desde os primeiros experimentos de expressão gênica utilizando *microarrays* de DNA (Lander 1999) e podem ser causados por muitos fatores, incluindo pequenas diferenças nos procedimentos experimentais, uso de diferentes lotes de *kits* de amplificação, ou mesmo oscilações nos níveis de ozônio atmosférico durante os experimentos (Fare et al. 2003; Branham et al. 2007). Estes efeitos são frequentemente observados quando novas amostras são adicionadas progressivamente a um conjunto de dados pré-existente, como neste projeto.

Alguns métodos já foram propostos para o ajuste de efeitos de lote (Benito et al. 2004; Alter and Golub 2006), mas frequentemente estes métodos requerem um número de amostras

elevado (>25) em cada lote e, portanto, não são adequados para o tratamento dos dados neste caso. Alternativamente, utilizamos o algoritmo chamado *ComBat* que corrige o efeito do lote utilizando o método Bayesiano empírico (*Empirical Bayes Method*). Este método foi proposto especificamente para correção de efeitos de lote em conjuntos de dados compostos por diversos lotes, com pequeno número de amostras em cada lote (Johnson et al. 2007), e foi utilizado com sucesso para a normalização de dados de expressão gênica obtidos ao longo do tempo, ou em diferentes plataformas (Abrahams et al. 2007; Coram et al. 2008; Martinez-Llordella et al. 2008), sendo capaz minimizar variações de processos não biológicos.

Na Tabela 2 são apresentadas as medidas da correlação média entre lâminas hibridizadas no mesmo lote ou em lotes diferentes. Utilizando-se os dados brutos (antes da normalização e filtragem), nota-se uma baixa correlação entre as lâminas dos lotes A e B (0,51), que se reflete em uma baixa correlação global (0,63). A normalização por quantil e filtragem dos dados pouco confiáveis contribuiram significativamente para aumentar a correlação intra lote, mas a correlação entre os 2 lotes permaneceu baixa (0,66). A correção do efeito de lote associado aos dados utilizado o programa *ComBat* tornou as correlações intra e inter lotes comparáveis (0,80), preparando os dados adequadamente para as análises subsequentes.

Tabela 2: Correlações médias entre medidas de expressão gênica obtidas das amostras hibridizadas nos dois lotes distintos.

		Dados Normalizados	
	Dados Brutos	sem correção de lote	com correção de lote
Correlação - Todas as lâminas	0.63	0.74	0.83
Correlação - lâminas Lote A	0.63	0.79	0.83
Correlação - lâminas Lote B	0.83	0.88	0.84
Correlação - lâminas Lote A vs.Lote B	0.51	0.66	0.80

Após a filtragem e normalização dos dados, seguida pelo ajuste entre os lotes pelo método empírico de Bayes, os dados de expressão gênica foram analisados para a

identificação de padrões de expressão gênica característicos de cada tipo histológico analisado. Para isso foi utilizada a abordagem SAM (*Significance Analysis of Microarray*) (Tusher et al. 2001), implementada como um *plugin* no programa Excel (Microsoft).

Optamos por utilizar a abordagem SAM para a identificação de assinaturas de expressão pois este método baseia-se na permutação dos dados de expressão para calcular uma taxa de descoberta de falsos-positivos (FDR) (Tusher et al. 2001; Storey and Tibshirani 2003) e dessa forma controlar o problema dos testes-múltiplos que afetam a identificação de genes diferencialmente expressos com significância estatística em experimentos de *microarrays* (Dudoit et al. 2004; Dupuy and Simon 2007). Uma limitação no uso de testes estatísticos clássicos como *teste-t* (comparações de 2 classes) e ANOVA (comparações de 3 ou mais classes) é que ambos os testes assumem que as medidas de expressão gênica possuem valores com uma distribuição normal, e que a variância é constante ao longo das diferentes faixas de intensidade de expressão gênica, o que não é necessariamente verdade em dados de microarranjos de DNA (Quackenbush 2002).

O modelo estatístico implementado no SAM utiliza um *teste-t* modificado, onde uma constante (*So*) é adicionada ao denominador do *teste-t*:

$$d(i) = \frac{\overline{X_A}(i) - \overline{X_B}(i)}{s_0 + s_i}$$

d(i): diferença relativa de expressão entre as amostras da condição A em relação à B;

$\overline{X_A}\ (i)\ e\ \overline{X_B}\ (i)$: valor médio de expressão do gene (i) na condição A e B;

s_i: desvio padrão das medidas de expressão das amostras;

s_0: constante positiva para minimizar o coeficiente de variação.

A constante é estimada através da soma global do erro-padrão dos genes e a adição deste valor reduz a chance de detecção de genes que possuem um pequeno desvio padrão, por apresentarem baixos valores de intensidade. O algoritmo favorece a atribuição de um alto valor para a constante *So*, o que remete que este teste estatístico fica mais dependente do valor de *fold-change* entre os grupos (Jeffery et al. 2006).

Para identificar genes com expressão correlacionada com a histologia do tecido pancreático foi utilizado o SAM com os seguintes parâmetros: análise *two-class*, 1000 permutações, *K-nearst Neighbors Inputer* e FDR (*False Discovery Rate*) ≤ 5% ou ≤10%. Em seguida, o programa *Spotfire* (SpotFire Inc., MA) foi utilizado para fazer o agrupamento hierárquico dos dados (*Ward's method*), de modo a agrupar genes e amostras com perfil de expressão semelhante.

3.7 Análise bioinformática dos transcritos intrônicos e intergênicos expressos nos tecidos pancreáticos.

Para realizar as análises bioinformáticas dos ncRNAs interrogados pelo microarranjo de DNA e detectados em tecidos pancreáticos, utilizamos o conjunto de programas de código aberto BEDTools (Quinlan and Hall 2010). Estes programas facilitam o cruzamento de arquivos com coordenadas genômicas de elementos de interesse (formato BED), como transcritos, genes, regiões promotoras, sítios de ligação de fatores de transcrição, entre outros, obtidos no *UCSC Genome Browser* (http://genome.ucsc.edu/). Com o uso desta ferramenta, foi realizada uma série de análises utilizando as coordenadas genômicas dos ESTs correspondentes a transcritos intrônicos, intergênicos e mRNAs depositados no *microarray* empregado nas hibridizações com amostras de tecido pancreático. As análises realizadas incluíram:

- Comparação do tamanho das ESTs utilizadas como sondas no microarranjo com a sequencia de ncRNAs preditos a partir de uma reconstrução bioinformática do transcriptoma;

- Avaliação da expressão dos ncRNAs interrogados no microarranjo em diferentes tecidos utilizando dados de RNA-Seq disponíveis publicamente;

- Avaliação da localização de ncRNAs intergênicos interrogados no microarranjo relativamente a regiões 3' ou 5' UTR de RNAs codificadores de proteínas conhecidos;

- Procura por elementos evolutivamente conservados nas sequências dos ncRNAs detectados em tecidos pancreáticos (conservação na sequência de DNA, expressão de mRNA e estrutura secundária).

- Busca por elementos regulatórios (H3K4me3 e CpG) na sequência de DNA dos lncRNAs estudados e busca de início de transcrição utilizando CAGE *tags*.

3.7.1 Comparação do tamanho das ESTs utilizadas como sondas no microarranjo com a sequencia de ncRNAs preditos a partir de uma reconstrução bioinformática do transcriptoma.

Para esta análise foram utilizadas sequencias de ncRNAs expressos no genoma humano obtidas a partir da montagem de 4.762.523 ESTs não redundantes, as quais se agrupavam em 332.946 *clusters* descritas no trabalho de Nakaya *et al.* (Nakaya et al. 2007). Cada cluster originou uma sequencia consenso denominada *contig*. Destes o autor identificou 55.139 *unspliced* ESTs *contigs* em regiões totalmente intrônicas (*Totally Intronic Noncoding RNA*, TIN) , sem sobreposição com exons de genes codificadores de proteína conhecidos. As coordenadas genômicas dos *contigs* gerados neste estudo encontram-se disponíveis no UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg17&position=chr1: 3112500-3152000&hgt.customText=http://genomebiology.com/imedia/1795171211288678/

supp13.psl). As coordenadas dos 55.139 *contigs* representando ncRNAs intrônicos referentes a montagem do genoma hg17 e foram convertidas para a versão mais atual (hg19) utilizando a ferramenta *lift-over* (http://genome.ucsc.edu/) usando os parâmetros padrão descritos pelo programa. A maior parte das coordenadas foi convertida com sucesso para hg19 resultando em 55.121 *contigs*. Após isto, as coordenadas genômicas das ESTs expressas em tecidos pancreáticos foram anotadas conforme descrito no item 3.5 e cruzadas com as coordenadas dos TINs utilizando a função do *BEDTool intersectBed* (Quinlan and Hall 2010) com parâmetro –u para retirar a redundância.

3.7.2 Avaliação da expressão dos ncRNAs interrogados no microarranjo em diferentes tecidos utilizando dados de RNA-Seq disponíveis publicamente.

Os dados de RNA-seq obtidos por Wang e colaboradores (Wang et al. 2008) e estavam depositados no site do UCSC (http://genome.ucsc.edu/). Como descrito pelo autor as bibliotecas não são direcionadas e foram geradas com *Illumina Genome Analyzer* (*Illumina*) a partir de RNA poliA(+) de nove tecidos humanos de doadores saudáveis (tecido adiposo, mama, cérebro, cólon, coração, fígado, linfonodo, músculo esquelético e testículo). As coordenadas referentes aos *reads* foram obtidas no UCSC em formato *BED* na montagem hg19 e a quantidade de *reads* mapeados em cada uma das bibliotecas utilizada está representada a seguir: Adiposo (17.635.339 *reads*), Cérebro (10.962.620 *reads*), Mama (10.557 .297 *reads*), Cólon (17.817.798 *reads*), Coração (10.630.949 *reads*), Fígado (11.571.688 *reads*), Linfonodo (16.221.077 *reads*), Músculo esquelético (14.212.800 *reads*) e Testículo (18.856.562 *reads*).

Em seguida, as coordenada das ESTs depositados no microarranjo foram cruzadas com as coordenadas dos *reads* de cada biblioteca de RNA-Seq utilizando a função do *BEDTool intersectBed*. Foi usado o parâmetro *–wo* para representar na saída do programa

todas as sobreposições de coordenadas. Assim, *reads* que continham *splicing* foram representados com duas coordenadas diferentes no arquivo *BED*, entretanto com o mesmo identificador. Para contar quantos *reads* se sobrepunham com cada *EST* foi escrito um programa em *perl*.

Para o cálculo de medidas de correlação entre a expressão de regiões intrônicas e exons adjacentes mostradas na Figura 12, as coordenadas genômicas de exons/introns flanqueando ncRNAs representados no microarranjo foram cruzadas com as coordenadas genômicas dos *reads* das nove bibliotecas de RNA-Seq. Os *reads* recuperados foram usados para estimar a abundância de cada região expressa, sob a forma de RPKM (*Reads per kilobase per million*) (Mortazavi et al. 2008).

$$RPKM = \frac{\text{N}^{\circ}\, reads}{N^{\circ}\ \text{total de}\ reads\ \text{da biblioteca}\ X\ \text{tamanho do transcrito}} \text{X } 10^{9}$$

Após a obtenção do RPKM de cada região (transcrito intrônico, exon à jusante e exon à montante), as correlações de Pearson foram calculadas e a distribuição mostrada no gráfico de frequência cumulativa.

3.7.3 Avaliação da localização de ncRNAs intergênicos interrogados no microarranjo relativamente a regiões 3’ ou 5’ UTR de RNAs codificadores de proteínas conhecidos

Para verificar se os ncRNAs intergênicos detectados no microarranjo poderiam ser partes ainda não anotadas de regiões 3’ ou 5’ UTR de RNAs codificadores de proteínas, as coordenadas referentes aos mRNAs conhecidos (conjunto de 40.789 mRNAs RefSeqs mapeados no hg19 – Setembro de 2011) foram obtidas nos site UCSC (http://genome.ucsc.edu/). As coordenadas referentes a esses mRNAs foram cruzadas com os

73 ESTs intergênicos usando a função do *BEDTool closestBed –d* (para calcular a distância mais próxima entre as coordenadas).

3.7.4 Procura por elementos conservados no DNA nas sequências dos RNAs expressos em tecidos pancreáticos

Os elementos conservados são representações de pequenas regiões discretas de DNA com alto nível de conservação entre as 46 espécies. Os elementos de conservação *PhastConsElements* 46way foram obtidos no UCSC (http://genome.ucsc.edu/), as espécies contidas no alinhamento múltiplo estão descritas na página do UCSC.

Foram utilizados elementos conservados em vertebrados, placentários e primatas. O número de elementos conservados em cada grupo está distribuído da seguinte forma: Vertebrados (n = 5.163.775, cobertura c.a. 4.6% do genoma humano), Placentários (n = 3.743.478, cobertura de c.a. 3.7% do genoma humano) e Primatas (n = 725.627, cobertura de c.a. 3.4% do genoma humano). As coordenadas dos ncRNAs representados no microarranjo foram cruzadas com as coordenadas genômicas obtidas em cada grupo usando a função *intersectBed* da ferramenta BEDTool (Quinlan and Hall 2010). O número de ncRNAs contendo elementos conservados foi dividido pelo número de elementos conservados em cada grupo (vertebrados, placentários e primatas) para que as medidas se tornassem comparáveis. Como controle foi utilizado um conjunto aleatório de sequências com mesmo conteúdo CG e mesmo tamanho em relação aos ESTs representados no microarranjo. O teste estatístico de *Fisher* foi utilizado para comparar as frequências de elementos conservados obtidas para os lncRNAs com as frequências de elementos conservados obtidas para as sequências randômicas.

3.7.5 Predição de estrutura secundária e potencial codificador

As análises de predição potencial codificador de proteína e de estrutura secundária e em larga escala foram realizadas utilizando os programas CPC (*Coding Potential Calculator*) (Kong et al. 2007) e RNAz (Gruber et al. 2007), respectivamente, rodados em linha de comando no ambiente Linux. O programa CPC calcula o potencial codificador de proteína de sequências de DNA/RNA analisando seis características intrínsecas da sequência. Primeiro, procura uma ORF (*Open Reading Frame*) predita usando o *framefinder* software. Após isso, se há identificação de uma ORF, o programa procura analisar a qualidade desta ORF, calculando um *score* (LOG-ODDS SCORE) que leve em consideração o tamanho e a integridade desta ORF. A integridade de uma ORF consiste em avaliar se esta começa com um *start codon* e termina com um *stop codon in frame*. Segundo o trabalho de Dinger e colaboradores (Dinger et al. 2008), uma das estratégias para diferenciar codificadores de proteína de ncRNAs é olhar o tamanho da *ORF*, uma vez que os lncRNAs são transcritos longos é possível que sejam identificadas *ORFs* pequenas em suas sequências. Uma estimativa de corte bom seria de 300 nt (100 aa) utilizado como *cutoff* no projeto *FANTOM* (Okazaki et al. 2002). Apesar deste número ser arbitrário, este está baseado na evidência de que mais de 95% das proteínas depositadas em bancos de dados públicos *Swiss-Prot* e *International Protein Index* possuem mais de 100 aa (Frith et al. 2006). Quanto maior a qualidade, levando em consideração o tamanho, e a integridade da ORF maior é a pontuação do *score* indicando ser maior a probabilidade da ORF predita codificar de fato um polipeptídeo. O programa ainda realiza um BLASTX para verificar se a ORF predita possui alguma similaridade com proteínas conhecidas depositadas no banco de dados do *UniRef 90* (UniProt Reference Clusters (Wu et al. 2006)). Neste caso, se a ORF possuir 90% de identidade com proteína conhecida calcula-se um HIT SCORE que mede a qualidade geral dos *hits*, que foram encontrados pelo BLASTX, considerando o número de hits encontrados e

qualidade desses *hits* (*E-value*). O programa ainda calcula um *score* denominado FRAME SCORE que avalia a distribuição dos hits encontrados pelo BLASTX nos 3 *frames* das ORFs, pois é possível encontrar ORFs preditas em transcritos não codificadores ao acaso, entretanto quando isto ocorre os *hits* tendem a estar espalhados nos 3 *frames* das ORFs, enquanto que para os transcritos codificadores de proteína, os *hits* encontrados pelo BLASTX estão, em geral, ancorados em um mesmo *frame*. Finalmente, o programa CPC calcula uma pontuação denominada CODING POTENTIAL SCORE levando todos os cálculos acima em consideração. Desta forma, transcritos preditos como não codificadores possuem uma pontuação menor que zero e os codificadores de proteína possuem uma pontuação acima de zero. A pontuação que fica entre no intervalo de [-1; 0] e [0; 1] é predita como potencial não codificador fraco e potencial codificador fraco, respectivamente. Logo, quanto mais longe o CODING POTENTIAL SCORE calculado estiver do zero, maior é a confiabilidade na predição do potencial codificador/não codificador.

O programa RNAz determina a predição de estrutura secundária baseando-se em dois parâmetros importantes: conservação estrutural e estabilidade termodinâmica. Inicialmente, o programa constrói uma sequência consenso predita usando o algoritmo *RNAalifold* (Hofacker et al. 2002). Este algoritmo utiliza as informações termodinâmicas e filogenéticas no modelo de *folding* por meio de energia livre. O RNAz determina a conservação estrutural calculando a taxa de energia da conformação consenso em relação aos trechos não conservados da sequência de cada espécie (SCI - *Structure Conservation Index*). Mutações compensatórias que preservam a estrutura secundária são consideradas na pontuação no modelo de energia livre. Para determinar a establidade da sequência consenso, o programa calcula o *z-score* da energia de conformação. Este valor representa o número de desvios-padrões acima ou abaixo a estabilidade de uma dada sequência está em relação a uma sequência randômica de mesmo tamanho e composição de bases. Assim, utilizando o SCI e o *z-score* o programa consegue

predizer se a sequência possuiu ou não estrutura secundária estável por meior do P valor. Somente sequências que apresentavam P ≥ 0.5 foram consideradas como estruturas secundárias estáveis.

3.7.6 Transmap – análise de conservação de expressão

Os transmaps mapeiam todos os tipos de transcritos conhecidos (ex. cDNA, ESTs, mRNAs presentes nos bancos de dados RefSeq ou UCSC) entre espécies utilizando alinhamento sintênicos, os quais levam em cosideração a posição dos genes localizados no mesmo cromossomo. As coordenadas dos Transmaps de transcritos expressos estão mapeados na montagem do genoma humano (hg19) disponíveis no site USCS (http://genome.ucsc.edu/) (Zhu et al. 2007). A metodologia de mapeamento utilizada para os Transmaps foi utilizar o alinhamento sintênico para achar ortólogos entre duas espécies usando BLASTZ (Schwartz et al. 2003). Para achar transcritos ortólogos expressos entre duas espécies, primeiro há o alinhamento de todos os mRNAs de uma espécie no próprio genoma, para depois realizar o alinhamento sintênico entre as duas espécies para que se possa projetar as coordenadas de uma espécie na outra. Para realizar essa análise haviam *transmaps* de 26 espécies disponíveis.

Para identificar os transcritos lncRNAs intrônicos e intergênicos que estavam expressos em outras espécis, as coordenadas dos ncRNAs foram cruzadas com as coordenadas do transmaps e o número de transcritos computados. Para verificar se o número encontrado de transmaps seria igual ao acaso foi utilizado um conjunto de sequências aleatórias de mesmo tamanho com a mesma distribuição cromossômica. O teste estatístico de *Fisher* foi utilizado para comparar as frequências de elementos conservados obtidas para os lncRNAs com as frequências de elementos conservados obtidas para as sequências randômicas.

3.7.7 Identificação de elementos regulatórios na vizinhança de ncRNAs

A identificação de elementos regulatórios em regiões do DNA vizinhas a loci contendo ncRNAs expressos nos tecidos pancreáticos foi realizada utilizando anotações de marcas de cromatina regulatórias e sítios de início de transcrição (*Transcription Start Sites*, TSSs) determinados experimentalmente e cujas coordenadas genômicas encontram-se disponíveis através de *tracks* do *Genome Browser* (http://genome.ucsc.edu/). Foram utilizadas coordenadas genômicas de marcas de trimetilação da lisina 4 da histona 3 (H3K4me3), ilhas CpG e CAGE *tags*, sabidamente associadas a regiões promotoras de genes eucarióticos. As análises de comparação de coordenadas genômicas foram realizadas utilizando o pacote de análise *BEDTools*.

As coordenadas genômicas das marcas de cromatina referentes à H3K4me3 foram medidas em 13 linhagens celulares (Mikkelsen et al. 2007), sendo que nenhuma delas referia-se ao tecido pancreático. Baseando-se no critério de escolha de Mikkelsen e colaboradores (Mikkelsen et al. 2007), somente picos que apresentavam p-valor $< 10^{-5}$ foram utilizados para limitar o ruído experimental, resultando em 471.010 regiões (64% do que havia sido representado inicialmente 731.844 regiões).

As ilhas CpGs utilizadas eram as ilhas canônica definidas por Gardiner-Garden e Frommer (Gardiner-Garden and Frommer 1987). Estas ilhas precisavam ter 50% ou mais de conteúdo GC, tamanho mínimo de 200 pb e razão de CG/(C+G) maior do que 0.6. Assim, incialmente, haviam apenas 28.691 regiões descritas.

Para análise de CAGE *tags*, as coordenadas também foram obtidas no site do UCSC (http://genome.ucsc.edu/), sendo os dados utilizados provenientes de 6 linhagens celulares. Os *tags* foram gerados a partir de RNA poliA(+). Todos os *tags* referentes a essas bibliotecas foram agrupados e formaram *contigs* maiores ou iguais ao tamanho dos CAGE *tags*. Somente

contigs que foram suportados por pelo menos 5 *tags* foram mantidos na análise resultando em aproximadamente 210.000 *contigs*.

Todas as marcas que se sobrepunham às regiões 5'UTR dos mRNA Ref-Seq foram excluídas para eliminar da análise marcas regulatórias associadas com genes já conhecidos. As coordenadas dos elementos estudados (H3K4me3, CpG e CAGE *tags*) mais próximas das extremidades das sequências dos ncRNAs intrônicos e intergênicos expressos nos tecidos pancreáticos foram selecionadas e a distância anotada. Como controle, foi utilizado um conjunto de 100 grupos compostos por sequências aleatórias com tamanho e conteúdo de CG similar das sequências não codificadoras analisadas. Para inferir um valor estatístico nas análises, foi utilizada a análise Kolmogorov-Smirnov, que mede a diferença de distribuição estatística entre dois grupos. Esta estratégia já foi utilizada por Mestdagh e colaboradores (Mestdagh et al. 2010). Para isso foi desenvolvido um *script* em R que utilizou o pacote Deducer (Team 2011) para comparar a distribuição de mapeamento dos elementos estudados em relação às extremidades das sequencias de ncRNAs e mRNAs codificadores de proteínas, com a distribuição destes elementos em relação às sequências dos 100 grupos aleatórios. Somente foram consideradas com significância estatística as distribuições que apresentaram p-valor < 0,05 em relação a todos os 100 grupos aleatórios.

3.8 Enriquecimento de categorias associadas a Gene Ontology (GO)

A análise de enriquecimento de categorias dos ncRNAs e mRNAs detectados em tecidos pancreáticos foi realizada por duas ferramentas bioinformáticas DAVID (*Database for Annotation, Visualization and Integrated Discovery*) e por IPA (*Ingenuity Pathway Analysis*).

A ferramenta bioinformática DAVID está disponível na internet (http://david.abcc.ncifcrf.gov/).A análise de enriquecimento se baseia nas proporções de

genes encontrados em cada categoria comparada à quantidade de genes encontradas no *background*. A comparação das proporções é realizada através do teste estatístico de *Fisher*, sendo o p-valor corrigido para testes múltiplos pelo *EASE Score* (Dennis et al. 2003; Hosack et al. 2003; Huang da et al. 2007). Esta correção de testes múltiplos EASE *Score* representa uma correção de testes múltiplos mais branda, onde classes pouco representadas no *background* são mais penalizadas do que classes menos representadas. Com o intuito de obter resultados mais robustos, foi utilizado como *background* em ambas as análises os *loci* gênicos (exons ou introns) representados no microarranjo que foram detectados nos tecidos pancreáticos (1607 *loci* no total). Apenas as categorias que possuíam um p-valor ≤ 0.05, corrigido por EASE *Score* foram utilizadas.

A *suite* de análise IPA foi utilizada para identificar redes moleculares enriquecidas em transcritos presentes nas assinaturas de expressão gênica associadas à cada tipo histológico. Os transcritos encontrados na assinatura como diferencialmente expressos com significância estatística são denominados "moléculas focais" nas redes encontradas pelo algoritmo do IPA. As redes são formadas pelas 3 categorias funcionais mais representadas entre todas as moléculas que as compõem, sendo que moléculas que não estavam presentes na análise (marcadas em branco) ou que não faziam parte dos transcritos selecionados na assinatura (marcadas em cinza) são consideradas para determinar as 3 categorias funcionais mais representadas. Tanto os genes que não faziam parte do *Background Dataset* ou que não compunham a assinatura foram inseridos na rede para conectar o maior número possível de moléculas focais. Para realizar a análise estatística, a ferramenta utiliza o teste de *Fisher* para calcular o p-valor das redes mais enriquecidas, comparando os transcritos que estão representados na assinatura com o conjunto inicial (*Background Dataset*) que são representados por todos os transcritos expressos da plataforma de *array*. O *score* calculado pelo IPA indica a probabilidade de encontrar a rede enriquecida ao acaso e corresponde a –log

do p-valor. Desta forma, quanto maior o valor do *score*, menor é a probabilidade do resultado obtido ser observado ao acaso.

3.9 RT-PCR fita-específica

Foram realizadas reações de RT-PCR fita-específica para determinar a orientação dos transcritos intrônicos detectados nos microarranjos. Amostras de RNA total de Mia PaCa–2 ou um *pool* de RNA de amostras clínicas de pâncreas foram utilizadas para a produção de cDNA utilizando o Kit SuperScriptIII *First-Strand Supermix* (Invitrogen). Duzentos nanogramas de RNA foram incubados a 65ºC por 5 minutos com 1µl de *primer* específico (2.5µM), 1 µl de tampão de anelamento 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 57ºC por 50 minutos, seguido de 5 minutos a 85ºC para inativação da enzima. Controles negativos sem *primer* e sem enzima foram feitos para verificar a ausência de auto-anelamento do RNA e contaminação de DNA genômico, respectivamente. Desta forma, para cada transcrito estudado foram realizadas 4 reações de transcrição reversa.

Os 20µl da reação anterior foram utilizados como molde para a PCR, em reações contendo 5µl de tampão 10X, 1.3µl de 50mM $MgCl_2$, 1µl de 10µM de dNTP, 1µl de *primer Forward* (10µM), 1µl de *primer Reverse* (10µM), 0.5µl de Taq Biolase e 20.2µl de água tratada com DEPC. A reação foi incubada a 95ºC por 5 min com 40 ciclos de PCR (95ºC por 30s; 55ºC por 30s; 72ºC por 30s) e 72ºC por 7 minutos.

3.10 RT-PCR quantitativo em tempo real

Para validar a diferença de expressão dos transcritos intrônicos foram realizadas reações de RT-PCR quantitativas em tempo real. Amostras de RNA de tumores primários (201T, 206T, 210T) e metástase (214M, 304M, 302M) foram utilizadas para a produção de cDNA utilizando o *Kit SuperScriptIII Frist-Strand Supermix* (Invitrogen). 1µg de RNA de cada amostra foi incubado a 65ºC por 5 min com 1µl de *primers randon hexamer* (50ng/µl), 1µl de 1 µl de tampão de anelamento 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 57ºC por 50 minutos, seguido de 5 minutos a 85ºC para inativação da enzima. Controles sem transcrição reversa foram feitos para verificar a ausência de contaminação por DNA genômico.

Os 20µl da reação de RT foram diluídos 1:7 com água tratada com DEPC, totalizando 160µl. Para cada transcrito foram realizadas reações em triplicata de PCR quantitativo em tempo real, utilizando 5µl de cDNA produzido juntamente com 5µl de *primers Foward+ Reverse* (4µM) referente a cada transcrito e 10µl de SYBER Green PCR Master *Mix* (Apllied BioSystem). A reação foi incubada a 95ºC por 20s e 40 ciclos de 95ºC por 3s, para desnaturação e 60ºC por 30s para o anelamento e extensão.

Os experimentos foram realizados em triplicata. O gene *HMBS* (Vandesompele et al. 2002) foi usado como controle endógeno para normalização dos níveis de expressão entre as amostra (Pfaffl 2001). Comparações entre os dois grupos de amostras foram representadas por *Fold Change* calculado em relação à amostra de menor expressão. O teste-t bicaudal foi utilizado para determinar a significância estatística da diferença de expressão, sendo apenas p-valores < 0.05 considerados como significativos. A tabela com os *primers* utilizados se encontra em Apêndice 1.

3.11 Clonagem de ncRNAs através de RACE.

Os experimentos de RACE (*Rapid Amplification of cDNA Ends*) foram realizados utilizando uma biblioteca de cDNA *ready-to-use* preparada a partir de tecido pancreático humano (MARATHON, Clontech). Esta biblioteca foi montada a partir de *primers* Oligo dT_{30} acoplado a uma sequência adaptadora divida em 2 partes AP1 e AP2, sendo que esta sequência também foi ligada na ponta 5' do transcrito após a síntese da primeira fita de cDNA. Os *primers* utilizados estão mostrados na Tabela 2. A amplificação dos transcritos por PCR foi feita utilizando o *primer* AP1. O produto do PCR foi diluído 50 vezes para a segunda reação de PCR com *primers* internos AP2 (*Nested PCR*). Todas as reações de PCR foram realizadas seguindo etapas de 5 ciclos a 94°C por 30s e 72ºC por 4min, seguidos de 5 ciclos a 94ºC por 30s e 70ºC por 4min e 25 ciclos a 94ºC por 30s por 4min. A tabela de *primers* utilizados se encontra em Apêndice 1.

Os produtos da segunda reação de PCR foram separados por eletroforese em gel de agarose a 1,5% e as bandas foram cortadas e extraídas seguindo o protocolo do *Kit Gel Extraction* da QIAGEN. As bandas foram clonadas no vetor pGEM (Promega) segundo as recomendações do fabricante em bactérias eletrocompetentes DH10B. Os plasmídeos foram extraídos segundo o protocolo de *Miniprep* (QIAGEN) e sequenciados. Como os *primers* utilizados para o sequenciamento continham sequências associadas ao vetor, inicialmente foi utilizado o programa *crossmatch* (Green and Ewing 2002) para mascarar a sequência do vetor e dos *primers* adaptadores. Em seguida, foi utilizado o programa *Phrap* para o alinhamento das sequências e montagem de *contigs* representando as sequencias obtidas por RACE.

3.12 *Northern Blot.*

Inicialmente foram realizados ensaios de *Northern blot* com sondas não radioativas utilizando o kit *BrightStar Psoralen-biotin Kit* (Promega). Entretanto, os resultados não se mostraram satisfatórios quanto a sensibilidade.Na tentativa de aumentar a sensibilidade foi testada a utilização de RNA enriquecido com poliA(+) e a marcação radioativa das sondas. Para isso, um total de 20ug de mRNA poliA+ células Mia PaCa 2 total foi aplicado em gel de agarose 1,2% contendo 2,2M de formaldeído. A extração do mRNA foi realizada partindo de aproximadamente 1,5mg de RNA total utilizando o kit *FastTrack MAG mRNA isolation kit* (Invitrogen), seguindo recomendações do fabricante. A transferência destes RNAs para membrana de *nylon* carregada positivamente (Ambion) foi feita por capilaridade. Foram usadas 3 sondas de oligonucleotídeos (50-60 nucleotídeos, 15 pmol, Tabela 3) marcadas radioativamente ([γ-32P]ATP (at 3,000Ci/mmol, 10mCi/ml, 50pmol total) através de reação com a enzima *T4 Polynucleotide Kinase* (10u) a 37°C durante 1h. Após a reação as sondas foram purificadas em colunas Amicon Ultra 0.5 mL (Millipore) e o volume total recuperado de cada sonda foi utilizado para a hibridização. As membranas foram pré-hibridizadas por 1h a 42°C em na solução ULTRAhyb-Oligo (Ambion) acrescido de DNA de esperma de salmão (100ug/mL) e depois hibridizadas durante 16h a 42°C. Após a hibridização as membranas foram lavadas por 2 vezes de com 2x SSC, 0.1% SDS e em seguida foi exposta a telas sensíveis a radiação durante 2 dias. Os sinais foram detectados através da leitura das telas no equipamento Typhoon™ 9400 (GE Healthcare), de acordo com as instruções do fabricante.

Tabela 3: Sequência de oligonucleotídeos utilizados no experimento de *Northern Blot*

Oligo	***Sequencia***
PPP3CB_antisenso 1	GAAATGTAGAAGTGTGCCTTGAGTTGAAAAATAAGTAGGAAAGTACAGGTCTTTAGA
PPP3CB_antisenso 2	CTGTGGCATGGAGTCAGTTTATACTTTCTTTTTCAGGTGAATACATTATAGTTCTTC
PPP3CB_antisenso 3	GGTCTCATAGGTGGATTTTTAGGAAGTTCTATGTTTTGAACAGAATTCAGCAATTAGTGG

3.13 Ensaio de promotor in vitro utilizando um gene repórter

3.13.1 Busca por predição de regiões promotoras *in silico*

Para fortalecer as evidências que o transcrito intrônico não codificador *DAPK-1* é de fato um transcrito independente, decidiu-se procurar por predições de promotores a montante do possível inicio de transcrição e marcas de cromatina que poderiam estar associadas à transcrição independente desse transcrito. A busca por regiões promotoras foi realizada utilizando o programa TSSG, disponível no site *Softberry* (http://linux1.softberry.com/berry.phtml?topic=tssg&group=programs&subgroup=promoter). Esta mesma região foi analisada utilizando o site *Alibaba2.1* (http://www.gene-regulation.com/pub/programs/alibaba2 /index.html) para procurar por possíveis fatores de transcrição que poderiam se ligar na sequência predita como promotora. Esta ferramenta prediz sítios de ligação de fatores de transcrição utilizando a base de dados *TRANSFAC* (Matys et al. 2006).

Além destas análises, também foi feita uma busca para identificar marcas de cromatina H3K4me3 associadas a regiões promotoras (Guttman et al. ; Mikkelsen et al. 2007). Para isso foram utilizados dados de marcas de cromatina disponíveis no UCSC (http://genome.ucsc.edu/). Foram analisadas 14 linhagens celulares, nenhuma relacionada ao tecido pancreático (Figura 5).

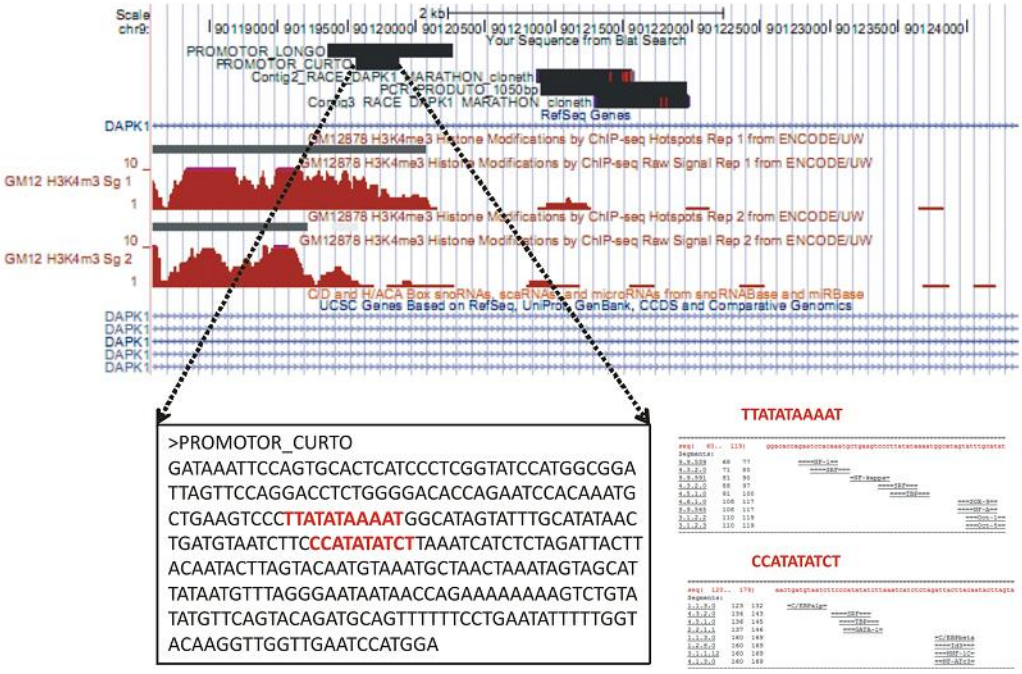

Figura 5: Diagrama representativo da região promotora do transcrito intrônico antissenso *DAPK1*. A região promotora do transcrito *DAPK1* se localiza a montante da região identificada pelo ensaio de RACE e RT-PCR fita-específica. A região determinada como promotor longo incluiu as duas regiões de ligação do TBP (*TATA-binding protein*) mais o início de marcação de H3K4me3 identificada na célula GM12878. A região denominada promotor curto apenas possui a região de ligação das proteínas TBP, que estão identificadas no painel abaixo marcadas em vermelho, e, ao, lado o resultado encontrado no site *alibaba2.0*.

3.13.2 Clonagem das possíveis regiões promotoras no vetor pGL3 (Promega)

Após essas análises *in silico* foi realizada a clonagem das possíveis regiões promotoras preditas para o ncRNA intrônico DAPK-1 (longa e curta) no vetor pGL3 (Promega). Como molde para essas reações foi utilizado DNA genômico da linhagem celular Mia PaCa-2. As

possíveis regiões promotoras foram amplificadas utilizando os pares de *primers* nas suas respectivas temperaturas de anelamento (termociclador MJ Research), usando a BIOLASE™ DNA Polymerase (Bioline) (Figura 6).

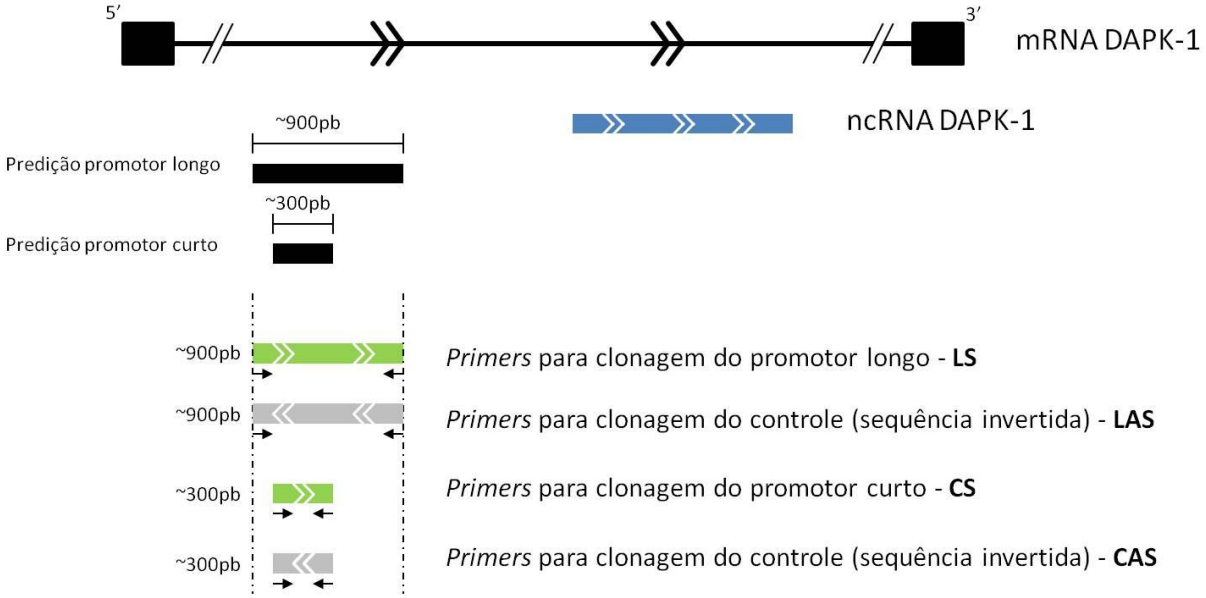

Figura 6: *Primers* utilizados para amplificação de possíveis regiões promotoras do lncRNA intrônico *DAPK-1*. Localização genômica dos diferentes segmentos usadas nas construções. As setas em preto mostram a região de anelamento dos *primers* que foram usados nas clonagens. As setas no interior das barras indicam a orientação dos fragmentos clonados.

As PCRs foram executadas em um volume final de 50 µl, contendo tampão 10x NH4-based *Reaction Buffer*, $MgCl_2$ 50 mM, mistura de dNTPs a 10 mM, 20 µM de cada oligonucleotídeo e 2,5 unidades de GoTaq Polymerase, completando-se o volume com água DEPC. Foram realizados 40 ciclos de desnaturação a 95°C por 1 minuto, anelamento em temperatura de 55°C por 30 segundos, e extensão a 72°C por 2 minutos. Os produtos de amplificação foram analisados em gel de agarose (1,2%) (Figura 7).

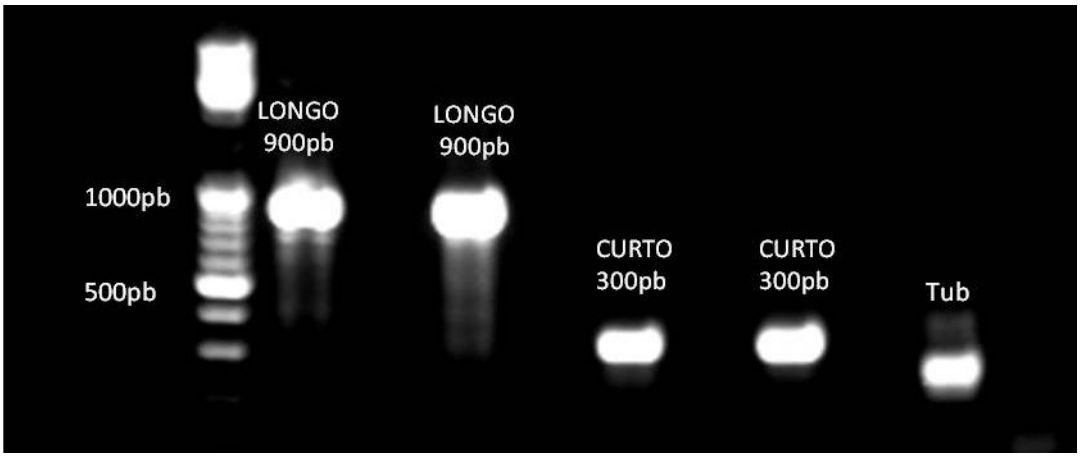

Figura 7: PCR a partir do DNA de MIA PaCa-2. Esta figura apresenta os tamanhos confirmados pelos *primers* desenhados para clonagem dos promotores curtos e longos. Esses fragmentos foram utilizados para a clonagem no vetor pGL3 para confirmar a presença de atividade promotora.

As bandas do gel de agarose foram purificadas usando o kit *Gel extraction* (Qiagen) e a quantidade de DNA obtida foi medida no espectofotômetro *Nanodrop ND-1000 (NanoDrop Technologies).* Foi realizada a reação de digestão do DNA purificado bem como a do vetor pGL3 com as enzimas de restrição NheI e Hind III (Figura 8). Foi utilizado 1µg de inserto, 10µl de cada enzima de restrição, 10µl de BSA (100x) juntamente com o tampão 10x NEBuffer 4. A reação foi mantida a 37 ^{o}C por quatro horas. Depois, foi realizada a reação de ligação vetor pGL3 + inserto *overnight* a 4 ^{o}C.

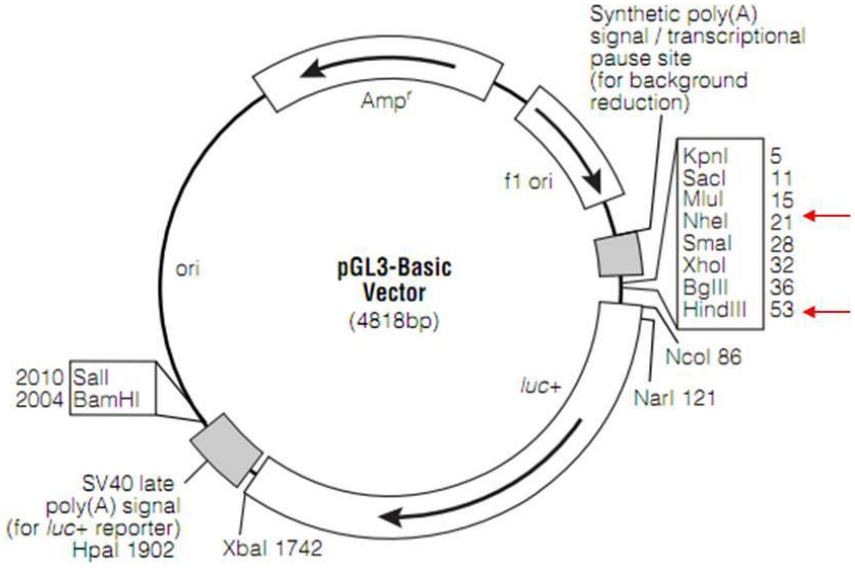

Figura 8: Representação esquemática do vetor usado para analisar a atividade r]promotora de uma posível região regulatória do lncRNA *DAPK-1* em linhagem celular MIA PaCa-2. Como controles também serão realizadas transfecções com o vetor vazio (sem nenhum inserto). As construções para os possíveis promotores (senso e antisenso) foram geradas a partir de uma estratégia de clonagem utilizando sítios específicos de enzimas de restrição (*NheI* e *HindI*, setas vermelhas).

Bactérias eletro-competentes da cepa DH10B foram transformadas com o produto da ligação por eletroporação no *Gene Pulser II* (BioRad), utilizando os parâmetros padrão. As bactérias transformadas foram cultivadas em placas de cultura contendo meio LB com ampicilina, IPTG e X-Gal, e incubadas em estufa a 37ºC *overnight*. Os plasmídeos recombinantes foram extraídos utilizando o *kit Spin MiniPrep* (Qiagen) e quantificados no espectrofotômetro *Nanodrop ND-1000 (NanoDrop Technologies).* Para confirmação da presença do inserto foi realizada a reação de digestão dos plasmídeos purificado com as enzimas de restrição NheI e Hind III (Figura 9).

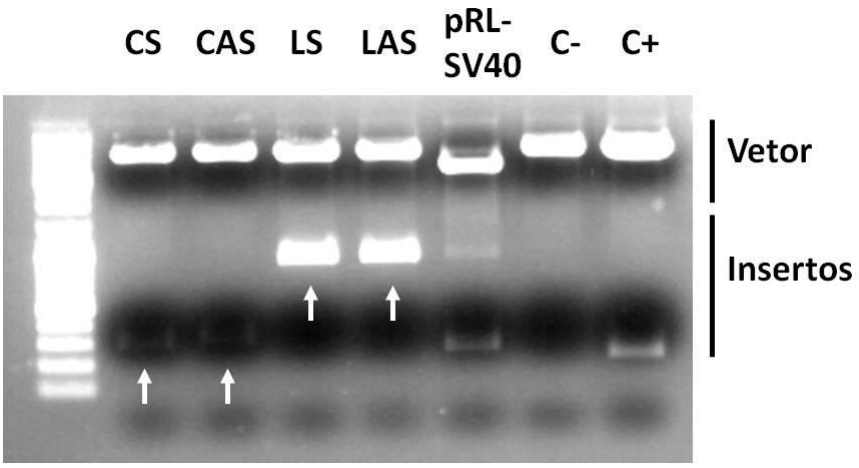

Figura 9: Confirmação da presença do inserto nos tamanhos esperados no vetor pGL3. Os plasmídeos foram digeridos com as mesmas enzimas de restrição utilizadas para fazer a construção, para confirmar a presença do inserto (indicado por setas). CS= promotor curto com orientação senso; CAS= promotor curto senso com orientação invertida; LS= promotor longo com orientação senso; LAS= promotor longo senso com orientação invertida; SV40= vetor controle pRL-SV40-control que possui o promotor forte SV-40 e Renilla Luciferase (controle transfecção); C- = vetor vazio pGL3-basic (controle negativo); C+ = controle positivo pGL3-control que possui promotor forte SV-40 e Firefly Luciferase.

Esse experimento confirmou a presença dos possíveis promotores clonados em pGL3 (CS, CAS, LS e LAS). Uma vez confirmada a presença dos possíveis promotores clonados em pGL3, esses foram transfectados em células Mia PaCa-2, como descrito a seguir.

3.14 Transfecção dos vetores pGL3 em células Mia PaCa-2

Foram plaqueadas $5x10^4$ células Mia PaCa-2 em placas de 24 poços, para que no dia da transfecção os poços apresentassem uma confluência de ~50%. Foram plaqueadas três réplicas biológicas de células que receberão os vetores carregando o possível promotor longo senso, longo senso invertido, curto senso, e curto senso invertido. Além disso, foram cultivadas também três réplicas biológicas células que receberão o vetor vazio pGL3-basic (controle negativo) e o vetor controle pGL3-control que possui o promotor forte SV-40

(controle positivo, *Firefly Luciferase*). O vetor controle pRL-SV40-control (controle de transfecção, *Renilla Luciferase*) foi também transfectado juntamente com células que receberão os vetores carregando o possível promotor longo senso, longo senso invertido, curto senso, curto senso invertido e vazio, para que seja realizada a normalização de cada medição comparando a atividade do possível promotor com a atividade do pRL-SV-40-control, uma vez que a eficiência da transfecção pode ser diferente entre as réplicas analisadas.

Para a transfecção, foram inicialmente adicionados à placa de cultura 1µg de plasmídeos, e em seguida foi adicionado 3,5 µL de Fugene. Essa mistura foi mantida a temperatura ambiente por 15 minutos e depois guardada em estufa por 48h.

3.15 Cultivo de Células Mia PaCa-2, PANC-1, AsPC-1 e Capan-1

Como o principal foco do trabalho era estudar lncRNAs intrônicos associados à patologia de tumor primário e metástase pancreáticos, foram selecionadas células humanas provenientes de tumores pancreáticos com e sem evidência de metástase. MIA PaCa-2 (ATCC CRL-1420) originada de carcinoma pancreático, sem evidência de metástase. PANC-1 (ATCC CRL-1469) célula epitelial originada de carcinoma do ducto pancreático. AsPC-1 (ATCC CRL-1682), célula de origem pancreática derivada de sítio metastático no fluido ascítico acumulado no peritônio. Capan-1 (ATCC HTB-79), célula de origem pancreática derivadas de metastático no fígado.

As linhagens celulares foram cultivadas de acordo com as recomendações da *American Type Culture Collection* - ATCC (http://www.atcc.org/). em frascos Corning 75cm^2 ou Nunclon 270mL para cultura de célula com 5mL ou 10mL. respectivamente. As linhagens Mia PaCa-2 e PANC-1 foram mantidas em meio de cultura DMEM composto em cada litro por glutamina e 1000mg/L glicose (D5523-1L Sigma-Alrich), 10mL de 100mM de piruvato de sódio (Gibco), 1.5 g/L de bicarbonato de sódio (Sigma), 0.04 g/L Penicilina (Cultilab), 0.1g/L Estreptavidina. As

linhagens metastáticas AsPC-1 e Capan-1 foram mantidas em meio de cultura RPMI composto em cada litro de e 4,5g/L glicose (D5523-1L Sigma-Alrich), 10mL de 100mM de piruvato de sódio (Gibco), 1.5 g/L de bicarbonato de sódio (Sigma), 0.04 g/L Penicilina (Cultilab), 0.1g/L Estreptavidina. Soro fetal bovino (SFB) foi adicionado às alíquotas de meio de cultura na proporção de 10% do volume total para cultivo de Mia PaCa-2, PANC-1 e AsPC-1 e 15% do volume total para o cultivo de Capan-1.

3.16 Silenciamento de transcritos utilizando oligos antisenso modificados

Para ensaios de silenciamento transiente dos transcritos selecionados foram adquiridos oligonucleotídeos antisenso modificados (IDT) com modificações *phosphotioate* e 2'O-metil, que são mais resistentes à degradação. A padronização dos experimentos de silenciamento foi realizada com MALAT-1, transcrito não codificador nuclear e actina, transcrito codificador de proteína. Ambos os transcritos são constitutivamente expressos em células Capan-1. Depois foram realizados experimentos de silenciamento do transcrito não codificadores *DAPK-S.*

Foram plaqueadas $5x10^5$ células Mia PaCa-2 em garrafas de 60mm com meio DMEM sem antibiótico e sem soro. Após 24hs, as células foram transfectadas com 120pmol dos oligos antissenso diluídos, inicialmente, em 3mL de OPTMEM-I (Gibco), utilizando 15µL de Lipofectamina RNAiMAX (Invitrogen). Após 4-6h, foram adicionados 3mL de meio DMEM com 20% de SFB sem antibiótico. Após 24 e 48hs, as células foram recolhidas e o RNA foi extraído com TRIzol e purificado com o RNeasy mini kit (QIAGEN) seguindo o protocolo do fabricante. Juntamente com a purificação, houve o tratamento do RNA com DNAse I para retirar eventual contaminação por DNA. Após o tratamento, foram realizadas reações de PCR

com *primers* para amplificar a tubulina a fim de confirmar a ausência de contaminação com DNA genômico nas amostras.

A síntese de cDNA foi realizada partindo de 1µg de RNA total utilizando o kit SuperScript®III First Strand Super*mix* (Invitrogen) seguindo as recomendações do fabricante. Após a síntese de cDNA, este foi diluído 8x (20µL de cDNA: 140µL de H_2O tratada com DEPC) para iniciar a reação de Real Time RT-PCR. Para a reação qPCR foram utilizados 5µL de cDNA diluído com 5µL de primer (4µM) e 10 µL de SyberGreen (Applied Biosystem), a reação foi realizada seguindo os parâmetros padrões recomendados pelo fabricante.

Os experimentos foram realizados em triplicata. A reação de *Real Time* RT-PCR foi realizada utilizando os *primers* relativos aos transcritos cujo silenciamento foram analisados e também com *primers* do gene normalizador *HMBS* (Vandesompele et al. 2002) e, portanto, nível de expressão foi calculado por quantificação relativa $\Delta\Delta C_T$ (Pfaffl 2001).

4 Resultados

4.1 RNAs não codificadores longos intrônicos e intergênicos são expressos em tecidos pancreáticos com diferentes graus de malignidade

Neste trabalho foi utilizado uma plataforma de microarranjos de cDNA customizada com aproximadamente 4000 elementos para investigar o padrão de expressão de RNAs codificadores de proteína e RNAs longos não codificadores de proteínas que mapeiam em regiões intrônicas e intergênicas do genoma em amostras clínicas de pacientes com câncer de pâncreas. Esta plataforma de microarranjos já foi descrita previamente em trabalhos do grupo [4, 5] e é composta por 2371 mRNAs de genes descritos na literatura como associados ao câncer, 910 transcritos mapeados em regiões intrônicas e intergênicas, além de 74 lncRNAs conhecidos. Os mRNAs e lncRNAs bem anotados fazem parte do conjunto RefSeq.

Alvos fluorescentes (cRNAs marcados com Cy5) foram gerados a partir de 38 amostras clínicas de pacientes contendo a seguinte distribuição: 15 amostras de adenocarcinoma de pâncreas primário, 6 amostras de metástases de tumores pancreáticos, 9 amostras de tecido pancreático não tumoral adjacente a tumores e 8 amostras de pancreatite crônica. Estas 38 amostras foram hibridadas individualmente com microarranjos e após a filtragem dos dados, 1607 transcritos foram identificados como expressos em pelo menos um dos tecidos utilizados (ver Materiais e Métodos, seção 3.6 para maiores detalhes). Estes incluíam 1267 mRNAs codificadores de proteína e 340 RNAs mapeados em regiões intrônicas e intergênicas (sem sobreposição com regiões exônicas de RefSeqs de RNAs codificadores de proteína), e 22 ncRNAs conhecidos como XIST (NR_001564.2), TSIX (NR_003255.2), MEG3 (NR_046465.2 entre outros RefSeqs). Somente transcritos cujas sondas no microarranjo apresentaram alinhamento com pelo menos 90% de identidade com RNAs anotados foram

selecionados para análises posteriores. Desta forma, cinco dos 78 transcritos intergênicos mostrados na Tabela 4 foram excluídos e um total de 335 ncRNAs (22 lncRNAs conhecidos, 240 possíveis lncRNAs mapeados em regiões intrônicas e 73 mapeados em regiões intergênicas) foram utilizados nas análises subsequentes.

Como pode ser visto na Tabela 4, mRNAs codificadores de proteína e os possíveis RNAs não codificadores longos mapeados em regiões intrônicas e intergênicas foram detectadas em todos os tipos histológicos de pâncreas. Cada tipo de RNA está representado de maneira comparável entre os tecidos pancreáticos utilizados.

Tabela 4: Distribuição de transcritos detectados em tecidos pancreáticos utilizando microarranjo com 4k elementos

Tipo	Nº sondas no microarranjo	Nº sondas detectadas como expressas em				Nº Sondas expressas *
		NT	T	M	PC	
mRNAs codificadores de proteína	2371	1106	1167	1198	1230	1267
lncRNAs conhecidos (RefSeq)	74	20	19	22	18	22
lncRNAs intrônicos	722	206	202	238	235	240
lncRNAs intergênicos	188	68	68	74	77	78
Total	3355	1400	1456	1532	1560	1607

*Para ser considerado expresso, o sinal da sonda deveria ser detectado acima da mediana da plataforma de microarrajo e estar presente em pelo menos 75% das amostras de pelo menos um tecido.

NT: tecido não tumoral, T: adenocarcinoma primário, M: metástase de tumores primários, PC: pancreatite crônica.

A fração de lncRNAs intrônicos detectados como expresso na plataforma de microarranjo foi de 0.33 (240/722), sendo comparável à fração dos lncRNAs conhecidos (22/74 = 0.30) e aos lncRNAs intergênicos (73/188 = 0.39). A pequena fração de lncRNAs detectado em tecidos pancreáticos (0.30-0.39) comparado com os mRNAs codificadores de

proteína (0.53) (Figura 10) reflete a observação de outros estudos que os RNAs não codificadores de proteína são menos abundantes e mais tecido-específico do que os transcritos codificadores de proteína (Birney et al. 2007; Nakaya et al. 2007; Sherman et al. 2007). De acordo com essa idéia, foi observado que a intensidade média dos possíveis lncRNAs detectados nas amostras de tecido pancreático foi menor do que a intensidade média dos mRNAs codificadores de proteína, com valores de 24,9 e 31,6 (unidades arbitrárias), respectivamente.

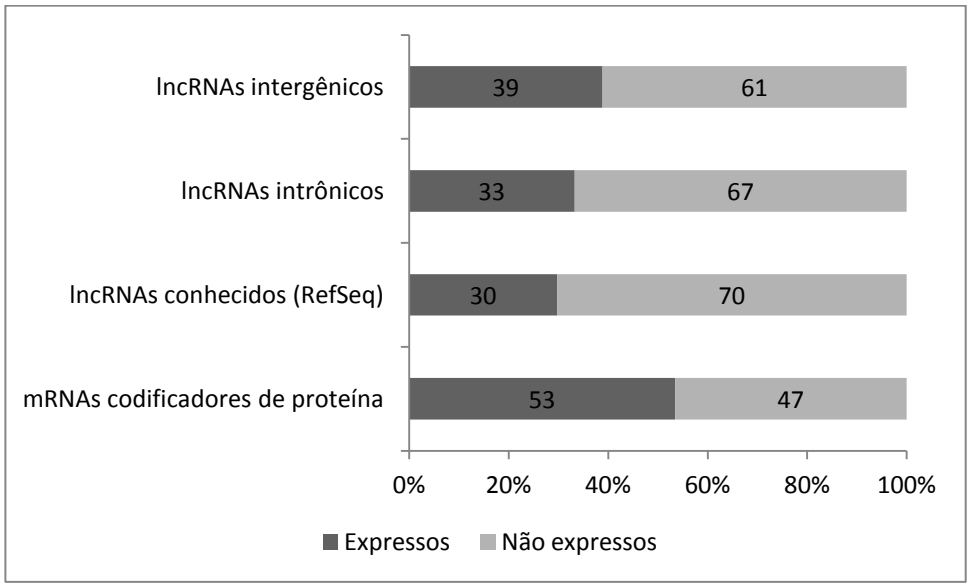

Figura 10: Fração de expressão de transcritos detectados em tecidos pancreáticos. O gráfico mostra em porcentagem (eixo X) a fração de transcritos detectados como expressos em pelo menos um dos 4 tecidos pancreáticos analisados em pelo menos 75% das amostras de um tipo histológico. No eixo Y está discriminado cada classe de transcritos estudados.

Dos 240 transcritos originados de regiões intrônicas identificados como expressos nos tecidos pancreáticos, apenas 62 possuíam sondas que interrogavam o transcrito codificador de proteína do mesmo *locus*. Desses, 31 (50%) apresentaram expressão apenas na região intrônica. Em 31 *loci* gênicos foram detectados simultâneamente transcritos intrônicos e transcritos codificadores de proteína. Para cada um desses *loci*, a correlação de Pearson (r)

referente à expressão do codificador de proteína e do transcrito intrônico foi calculada utilizando todas as amostras de tecido pancreático. Na maior parte dos casos (27 dos 31, 87%) foi observada uma baixa correlação (-0.5 < r < 0.5), com 12 *loci* apresentando correlação negativa entre a expressão do possível lncRNA e o mRNA codificador de proteína, e 19 mostrando a correlação positiva como pode ser visto na Figura 11.

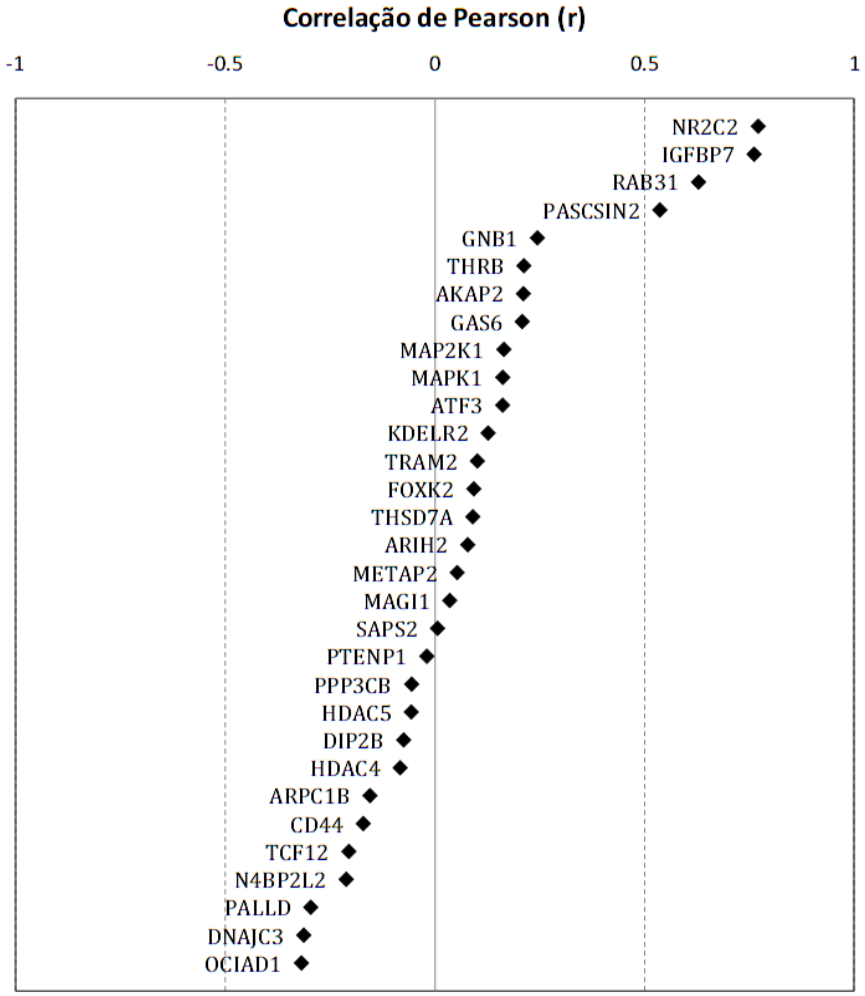

Figura 11: Cálculo da correlação de Pearson (r) referente ao nível de expressão do transcrito intrônico e o codificador de proteína. O gráfico acima mostra a correlação de Pearson (r) representada no eixo X. Cada ponto representa um *lócus* gênico onde foi calculada a correlação de expressão entre o transcrito intrônico e o transcrito codificador correspondente. No total foram calculados 31 coeficientes de correlação de Pearson e cada um está representado com o seu respectivo *Gene Symbol*. Somente 4 *loci* gênicos apresentam correlação positiva alta (13%, 4/31), a maior parte (87%, i.e. 27/31) apresenta baixa correlação (-0.5 < r < 0.5), sendo que desses 27, doze (1**2**/27, 4**4**%) apresentam correlação negativa (r < 0) e 15 (1**5**/27, **56**%) apresentam correlação positiva.

Dos 31 *loci* analisados, , treze (42%) apresentaram a mediana de expressão do transcrito intrônico maior do que a do transcrito codificador entre as 38 amostras clínicas de pacientes. Dos 13 transcritos intrônicos, 9 (69%) apresentaram mediana de expressão 2 vezes maior do que o seu respectivo codificador de proteína. Esses resultados sugerem que esses transcritos intrônicos transcritos sejam originados de uma transcrição independente, e não um sub-produto do processo *splicing* pós-transcricional do mRNA como apontado por alguns grupos (Struhl 2007; van Bakel et al. 2010).

Para estender essa análise de correlação para os outros transcritos intrônicos detectados como expressos na análise de tecido pancreáticos, foram utilizados dados públicos de RNA-seq de 9 tecidos diferentes (Wang et al. 2008) e a correlação de Pearson entre a expressão dos ncRNAs intrônicos e dos codificadores de proteínas correspondentes ao mesmo *locus* foi calculada. Para o cálculo de expressão foi utilizada a abordagem de RPKM (*Reads Per Kilobase of exon model per Million mapped reads*) (Mortazavi et al. 2008), que normaliza os dados de expressão pelo tamanho dos transcritos e tamanho da biblioteca sequenciada. Para o cálculo de correlação de Pearson, foram utilizados os níveis de expressão (RPKM) do transcrito intrônico, dos éxons que se situavam à montante e à jusante do transcrito intrônico. Somente os *loci* gênicos que apresentassem todos os elementos (região intrônica, éxon à montante e o éxon à ajusante) detectados em pelo menos das 4 das 9 bibliotecas de RNA-Seq disponíveis foram utilizados para o cálculo do coeficiente de Pearson. Foi verificado que 75/240 (31%) *loci* gênicos satisfizeram essa condição.

Como esperado, foi observada uma alta correlação entre a expressão dos éxons flanqueando regiões intrônicas, que possivelmente compõe o mesmo mRNA maduro (66/75, $r > 0.5$, Figura 12, linha preta). Com relação aos pares intron-éxon, um terço dos pares intron-éxon mostrou uma correlação positiva alta (24/75, $r > 0.5$). A expressão de mais da metade dos pares intron-éxon apresentou baixa correlação e uma pequena fração apresentou

correlação negativa alta (5/75, r < -0.5). A baixa correlação observada entre a abundância de transcritos intrônicos e de transcritos provenientes de éxons adjacentes sugere que a maior parte dos RNAs intrônicos sejam acumulados e processados na célula de forma distinta dos mRNAs produzidos no mesmo *locus*.

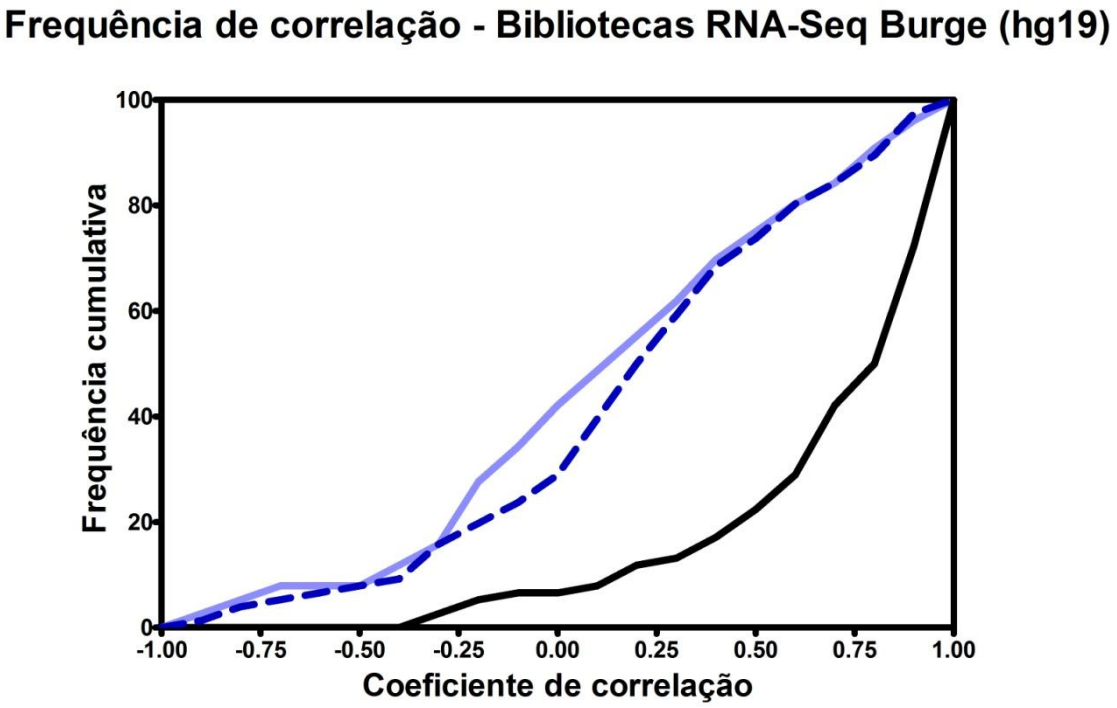

Figura 12: Cálculo de correlação entre regiões intrônicas e éxons adjacentes usando as 9 bibliotecas de RNA-seq (Burge RNA-seq). O número de *reads* foi normalizado por RPKM. Somente transcritos que apresentassem a expressão éxon-intron e éxon-éxon em pelo menos 4 bibliotecas tiveram a correlação de Pearson calculada. Utilizando este critério o número de transcritos utilizados nessa análise foi de **75** (**75/240, 31%**). Importante ressaltar que 50% dos coeficientes de correlação calculados entre as regiões exônicas (*upstream*/à montante e *downstream*/à jusante) foram iguais ou maiores a **0,8**, enquanto que 50% dos coeficientes de correlação entre as regiões intrônicas e exônicas adjacentes foram iguais ou menores a **0,2** para os **75 transcritos** analisados. Destes, metade apresentou uma correlação negativa, representando aproximadamente 25% do total. Essa baixa correlação entre os transcritos de regiões intrônicas e os éxons adjacentes sugere que parte destes transcritos intrônicos possam ser originados de transcrições independentes.

As mesmas bibliotecas de RNA-seq foram utilizadas para verificar o padrão de expressão dos 335 transcritos não codificadores detectados em tecidos pancreáticos nos nove tecidos disponíveis. Para realizar esta análise, as coordenadas genômicas dos ncRNAs detectados em tecidos pancreáticos foram cruzadas com as coordenadas genômicas dos *reads* provenientes de cada uma das nove bibliotecas de RNA-Seq. Foi observado que aproximadamente 80% dos transcritos detectados em tecidos pancreáticos são também expressos em pelo menos um dos nove tecidos humanos analisados (Figura 13).

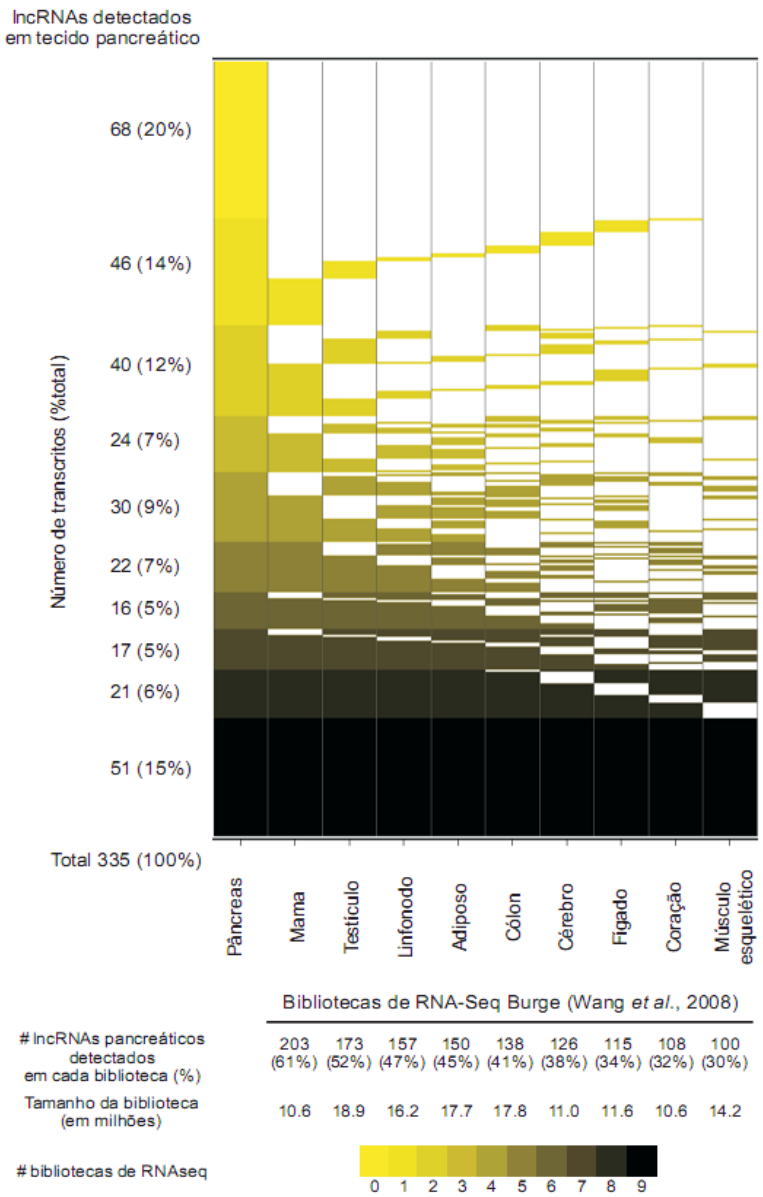

Figura 13: RNAs intrônicos e intergênicos expressos em tecidos pancreáticos são detectados em diferentes tecidos humanos. Coordenadas genômicas de 335 RNAs intrônicos/intergênicos detectados nos tecidos pancreáticos por microarranjo de DNA tiveram as coordenadas genômicas comparadas com 9 bibliotecas de diferentes tecidos geradas por RNA-Seq. Os ncRNAs foram agrupados de acordo com o número de bibliotecas identificadas (eixo Y). As bibliotecas de RNA-Seq estão apresentadas no eixo X e ordenadas, da esquerda para direita, de acordo com o número de ncRNAs pancreáticos detectados em cada biblioteca. Na coluna da esquerda estão representados o número/porcentagem dos transcritos expressos em pâncreas que também são identificados em outros tecidos. Na parte de baixo é apresentado o número de *reads* em cada biblioteca e o número de lncRNAs expresso em pâncreas também detectado em cada uma das bibliotecas. As cores representam o número de bibliotecas em que o transcrito foi identificado.

Como mostrado na Figura 13 acima, as maiores frações são representadas por transcritos que foram detectados apenas em tecidos pancreáticos ou transcritos detectados em todos os 9 tecidos analisados por RNA-Seq. Isto sugere que os ncRNAs são heterogêneos quanto ao padrão de expressão tecidual, incluindo transcritos com expressão restrita a tecidos específicos ou amplamente expressos em diferentes tecidos.

4.2 Análise *in silico* do potencial codificador, tamanho e processamento em pequenos RNAs dos transcritos intrônicos expressos em tecidos pancreáticos

Para caracterizar melhor os transcritos intrônicos e intergênicos quanto ao potencial de codificarem polipeptídeos, o seu potencial codificador foi medido utilizando o *software* CPC (*Coding Potential Calculator*) (Kong et al. 2007). Essa análise mostrou que a maior parte dos transcritos intrônicos e intergênicos (322/335, 96%) possui baixo ou nenhum potencial codificador. Basicamente, o *software* procura por ORFs na sequência das ESTs nas 6 fases de leitura e alinha as sequências nucleicas obtidas (BLASTX) contra o banco de proteínas *UniRef 90* (UniProt Reference Clusters, (Wu et al. 2006)) para verificar se a ORF predita possui alguma similaridade com proteínas conhecidas. Como pode ser notado na Figura 14 os

RNAs (intrônicos e intergênicos) apresentaram tamanho de ORFs e *score* CPC menores que os observados para mRNAs codificadores de proteína, indicando que a maior parte dos transcritos intrônicos e intergênicos, de fato, são RNAs não codificadores.

A Figura 14 mostra que a maior parte dos transcritos mapeados em regiões intrônicas e

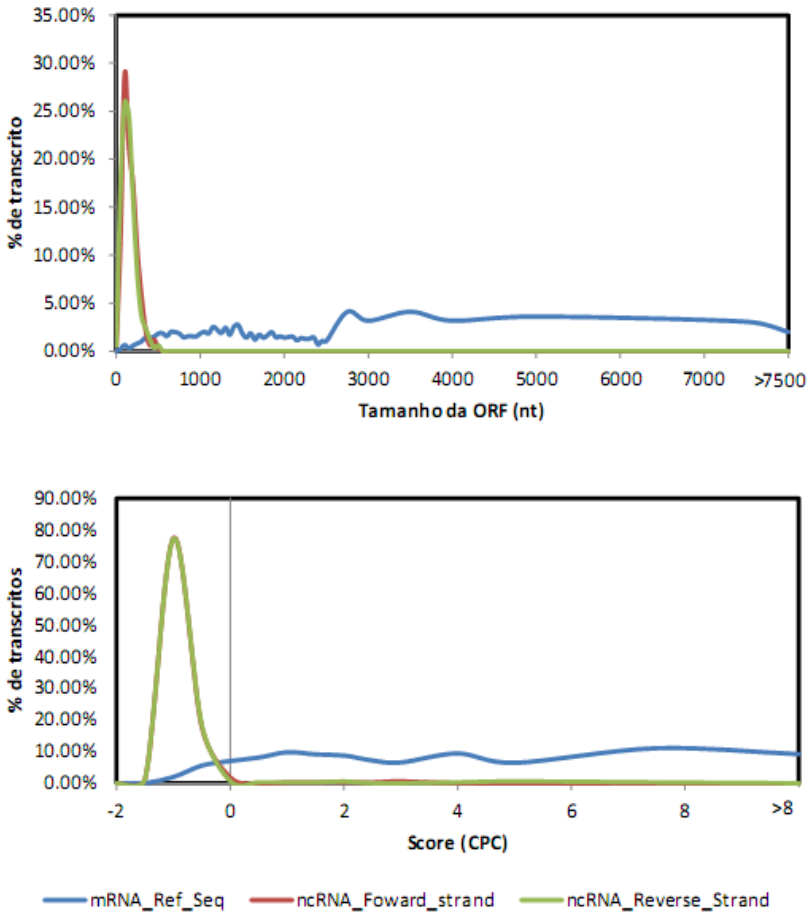

Figura 14 : As ORFs encontradas para os ncRNAs são curtas e com baico potencial codificador. O primeiro painel representa o tamanho das ORFs (eixo X) e a porcentagem de transcritos (eixo Y). Como pode ser visto no primeiro painel, a maior parte dos transcritos (95%) não codificadores possuem ORFs menores que 300nt. No segundo painel está representado o *Score* calculado pelo programa CPC (eixo X). Valores menores que zero (linha cinza) são classificam o transcrito como não codificador e valores maiores que zero classificam o transcrito como codificador.

intergênicas possui uma *ORF* pequena de aproximadamente 300nt, isto representaria uma proteína de 100aa, podendo representar um *ORF* confiável ou não. O programa também faz um BLASTX para calcular o número de *hits* encontrados em comparção à base de dados de proteína UniRef 90, pois é provável que transcritos codificadores de proteína possuam maior quantidade de *hits* do que os ncRNAs. Em geral, o tamanho pequeno da *ORF* já é um indicativo de baixa qualidade, entretanto há outros parâmetros que são levados em consideração como mostrados em na seção 3.7.5 em materiais e métodos, isso indica que as *ORFs* preditas para os RNAs intrônicos e intergênicos possuem grande chance de não serem confiáveis e, portanto, a maior parte desses expressos em tecidos pancreáticos pode ser classificada como não codificadores. Além disso, os ncRNAs apresentam um padrão de valor de *score* e tamanho de *ORFs* bem diferentes dos padrões apresentados pelos codificadores de proteína, indicando que se tratam de grupos bem distintos de RNAs. Ainda, a maior parte dos ncRNAs (intrônicos e intergênicos) possuem um *score* menor do que zero, o que os classifica como RNAs não codificadores de proteína.

Visto que os transcritos não codificadores longos são aqueles denominados, arbitrariamente, com aqueles RNAs não codificadores que possuem sequência de 200 ou mais nucleotídeos (Baker 2011), o tamanho das ESTs referentes aos transcritos intrônicos (n = 240) foi investigado. Primeiro, foi realizado um análise com o conjunto de sequências intrônicas a fim de verificar o tamanho das ESTs e estes transcritos foram comparados com o banco de TINs e PINs depositados no UCSC (http://genome.ucsc.edu/) que foram gerados por nosso próprio grupo por Nakaya e colaboradores (Nakaya et al. 2007). Estas denominações TINs e PINs foram originadas de contigs montados a partir de sobreposição de ESTs, sendo os transcritos TINs denominados como totalmente intrônico e PINs, parcialmente intrônicos. Desta forma, os transcritos TINs possuem uma montagem de *contigs* localizados apenas em regiões intrônicas e os transcritos PINs possuem montagem de *contigs* que passam por regiões

intrônicas e atravessam regiões exônicas, cobrindo totalmente o éxons possuindo, pelo menos, 30nt em ambas as pontas das regiões exônicas.

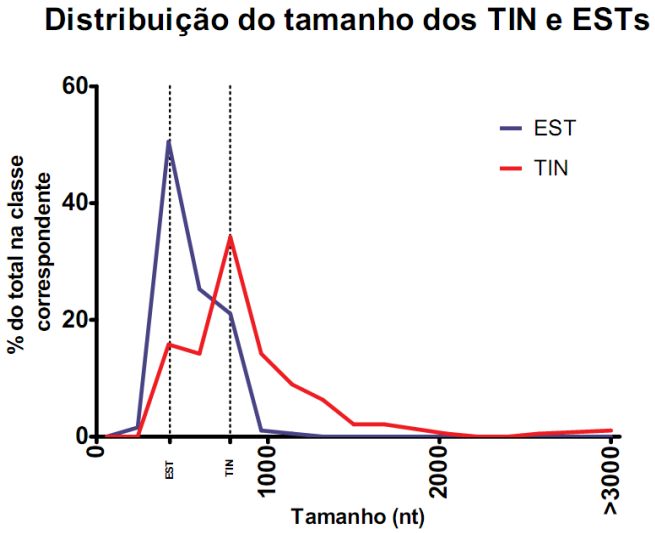

Figura 15: Distribuição de tamanhos das sequências de TIN e ESTs. O gráfico mostra a distribuição de tamanhos das sequências dos TIN contigs (vermelho) e ESTs (azul). O eixo Y representa a porcentagem das sequências. As marcas pontilhadas representadas por TIN e EST mostram o tamanho médio das sequências, sendo 779nt e 428nt, respectivamente.

Nesta análise, foi visto que 190 dos 240 (79%) transcritos intrônicos estavam localizados nas montagens de *contigs* TINs (Figura 15). O tamanho médio dos TINs encontrados foi de 779nt, enquanto que o tamanho médio representado pelas ESTs foi de 428nt. Comparando as 50 sequências restantes, foi visto que 11 sequências continham uma predição de PIN *contig*. O tamanho médio dos PIN *contigs* foi de 4129nt, no entanto o tamanho médio apresentado pelas ESTs que mapeavam em alguma predição PIN foi de 405nt. Este resultado sugere que as sequências de ESTs que estão representadas na plataforma de

microarranjos são sequências parciais de transcritos mais longos, logo estes transcritos analisados podem ser classificados como longos RNAs não codificadores (lncRNAs).

No caso dos transcritos intergênicos não foi possível realizar essa análise de sobreposição entre TINs e PINs pelo fato destes últimos estarem localizados dentro de coordenadas de genes conhecidos. Entretanto, para averiguar se os transcritos intergênicos pudessem ser possíveis UTRs não anotados ainda, foi realizada uma análise entre as coordenadas dos 73 transcritos intergênicos e dos UTRs mapeados em RefSeqs (montagem hg19, Outubro de 2010). Dos transcritos analisados 14 foram encontrados a pelo menos até 1kb de distância de 5' ou 3' UTRs conhecidos, sugerindo que 19% desses transcritos possam representar regiões UTRs de mRNAs codificadores de proteínas conhecidos ainda não anotados.

É bem estabelecido na literatura que pequenos RNAs regulatórios são originados do processamento de RNAs longos transcritos de regiões intrônicas ou intergênicas do genoma (Wilusz et al. 2009). Para verificar se os RNAs longos não-codificadores detectados em tecidos pancreáticos poderiam ser fontes de pequenos RNAs, o conjunto de 335 transcritos foi comparado com a base de dados de microRNAs e snoRNAs (Lestrade and Weber 2006; Griffiths-Jones et al. 2008). Nenhuma similaridade entre os transcritos analisados e as bases de dados utilizadas foi encontrada com exceção de uma sequência que mapeou no *locus SNORD89*. Entretanto, como é relatado na literatura que o tamanho médio de precursores de RNAs pequenos (>1Kb) (Bartel 2004; Du and Zamore 2005) é maior do que o tamanho médio apresentado pelas ESTs (428nt), a análise foi realizada novamente estendendo as coordenadas genômicas das sequências de ESTs com 1Kb em cada ponta do transcrito. Nesta nova análise foram encontradas mais 4 sequências que mostravam alta similaridade com 7 RNAs pequenos representados a seguir: *hsa-mir-1259, hsa-mir-326, hsa-mir-4269, hsa-mir-*

675, SNORD12 , SNORD12B e *SNORD12C*. A baixa sobreposição encontrada não exclui a possibilidade de que os ncRNAs longos intrônicos sejam precursores de pequenos RNAs ainda não anotados.

A fim de obter mais informações descritivas sobre esses transcritos, as coordenadas desses 335 lncRNAs foram cruzadas com a última montagem do GENCODE v12, disponível no site do UCSC (http://genome.ucsc.edu). Nesta análise, foram encontrados 96 transcritos (29%) que mapeiam em regiões de transcritos não codificadores definidos pelos GENCODE v12. Dos 96 transcritos, 27 mapeiam em regiões exônicas das predições de lncRNAs definidas por essa última montagem do GENCODE, sugerindo que esses lncRNAs intrônicos e intergênicos possam ser novos RNAs não codificadores ainda não anotados.

Ainda para a caracterização dos ncRNAs intrônicos, foi feita uma análise de sobreposição das coordenadas de mapeamento dessas sequências com regiões conservadas discretas (*conserved DNA elements*) anotadas nos genomas de Vertebrados, Mamíferos e Primatas. A conservação de sequência entre espécies evolutivamente relacionadas sugere a existência de seleção purificadora, que por sua vez é frequentemente associada como uma evidência de importância funcional para uma dada região genômica (Ponjavic et al. 2007). Nesta análise foi observado que apenas uma porção (n = 154, i.e. 46%) das sequências expressas nos tecidos pancreáticos se sobrepõe a regiões contendo elementos conservados. Foi observado também que a frequência de sobreposição é maior com o aumento da especificação do grupo (primatas > mamíferos > vertebrados). A sobreposição com elementos conservados observada para os ncRNAs foi maior do que a observada para o grupo controle, composto por sequências genômicas selecionadas ao acaso com mesmo tamanho e conteúdo de CG (*Fischer's exact test* p-valor < 0.05, Figura 16), indicando que a sobreposição com os transcritos não codificadores expressos em tecidos pancreáticos não ocorre ao acaso. Apesar de ser observado um enriquecimento de sequências conservadas nos ncRNAs expressos em

tecidos pancreáticos, esta sobreposição foi pequena possuindo uma taxa de no máximo 30% das sequências observadas quando comparadas com o nº total de ncRNAs expressos. Entretanto, a falta de conservação das sequências, que não possuíam sobreposição com elementos conservados, não indica a falta de função ou importância, pois como relatado na literatura, os RNAs não codificadores possuem uma baixa taxa de conservação na sua sequência global (Pang et al. 2006).

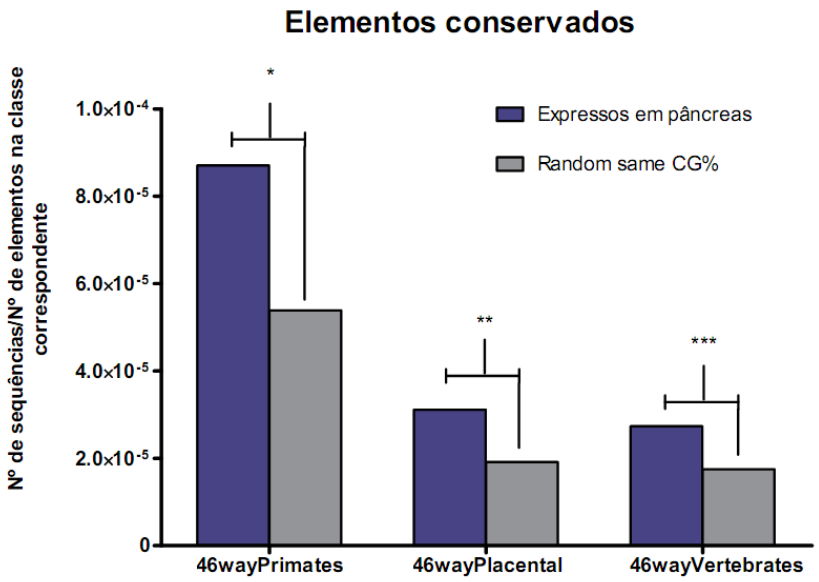

Figura 16: Sobreposição de ncRNA intrônicos com elementos de DNA conservados identificados na comparação com 3 diferentes grupos taxonômicos. Coordenadas de elementos de DNA conservados em Vertebrados, Mamíferos e Primatas (PhastCons depositados no UCSC foram cruzadas com as coordenadas de ncRNAs intrônicos expressos em tecidos pancreáticos (barras azuis). Como controle foram utilizadas coordenadas de sequencias intronicas aleatoriamente selecionadas no genoma com o mesmo conteúdo de GC (random same %CG, barras cinza). A sobreposição observada para os ncRNAs foi maior do que a observada para o grupo controle (*Fischer's exact test p-value* < 0.05), sugerindo que não ocorre ao acaso. O tamanho das barras indica o número de sequências encontradas em cada grupo dividido pelo número total de elementos conservados em cada grupo.

Outra análise realizada foi verificar se os transcritos não codificadores expressos em tecidos pancreáticos apresentavam estrutura secundária conservada utilizando o RNAz (Gruber et al. 2007). Como porposto por Washielt *et al*. (Washietl et al. 2005), a busca por estrutura conservadas pode levar a descoberta de novos RNAs não codificadores funcionais. Realizando esta análise, foi visto que 49 sequências apresentavam estrutura secundária conservada (15%), sugerindo que pelo menos uma porção desses RNAs contém domínios estruturais conservados que poderiam ser importantes para processamento e/ou para função biológica. As estruturas conservadas se encontram no site abaixo http://verjo102.iq.usp.br/sites/tahira/structures.html.

Além da busca de conservação de sequência de DNA e estrutura secundária, também foi analisado a conservação de expressão desses lncRNAs em outras espécies utilizando *transmaps*. A Figura 17 mostra que 100 transcritos lncRNAs (30%, 100/335) foram identificados como expressos em pelo menos 19 outras espécies. Apesar da taxa se apresentar baixa, esta ainda é estatiscamente significante quando comparada com o grupo randômico de mesmo tamanho e distribuição cromossômica (*Fischer's exact test* p-valor < 0.05). O baixo número de transcritos intrônicos e intergênicos com conservação de expressão pode ser explicado pelo fato da limitação de dados disponíveis referentes ao transcriptoma de várias espécies, pois grande parte do que foi encontrado foram originados de camundongo que apresenta uma base de dados bem extensa devido ao projeto FANTOM (*Functional Annotation of the Mouse Genome*), o qual se dedica a anotação funcional desta espécie (Okazaki et al. 2002). Espécies mais próximas como macaco *Rhesus* e orangotango possuem 62.206 e 46.026 ESTs depositadas no UCSC, enquanto que camundongo possui 4.370.322 ESTs, suportando o argumento acima.

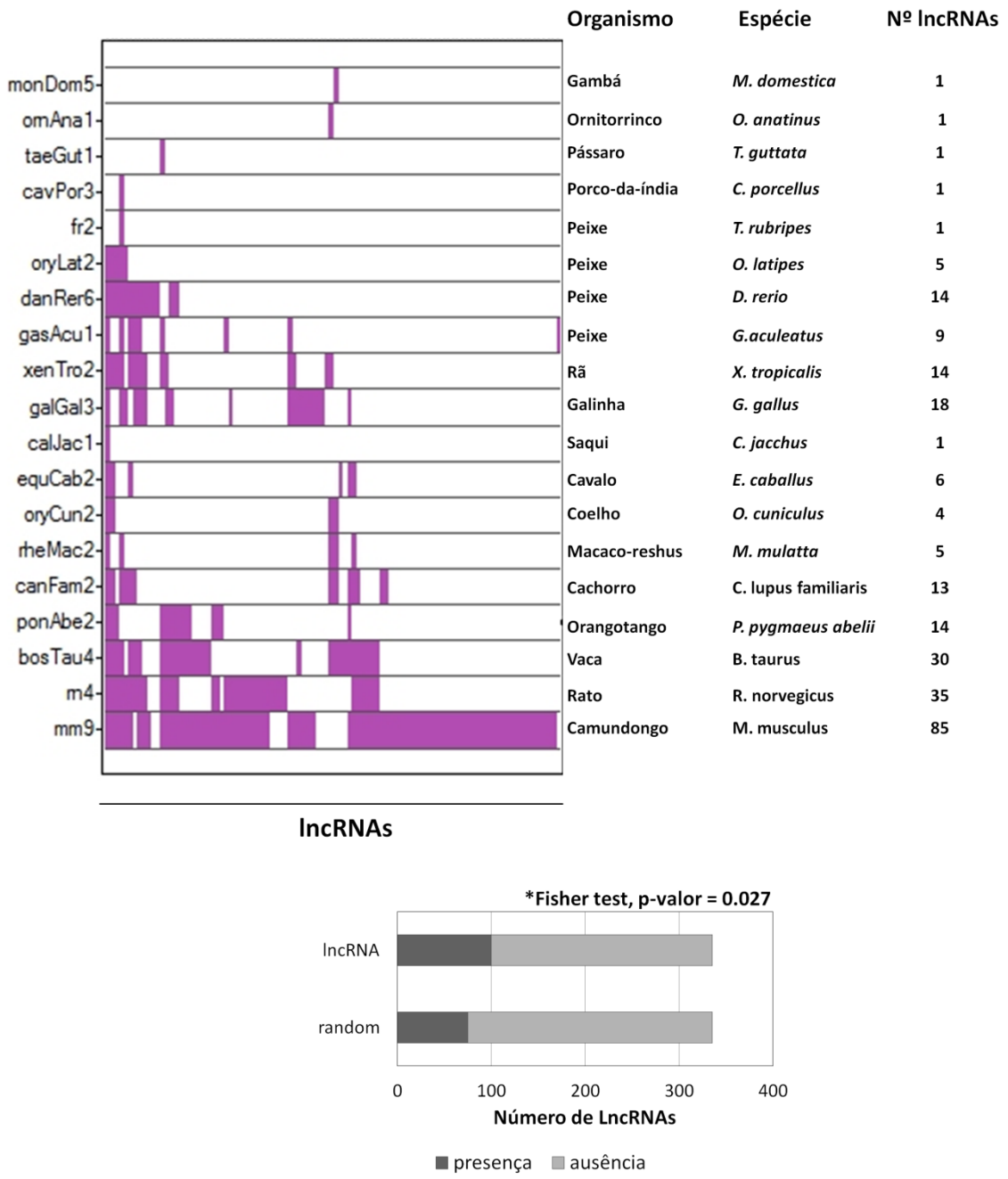

Figura 17: Análise de conservação de expressão e sequencia de ncRNAs intrônicos A matriz mostra lncRNAs intrônicos ou intergênicos (n=100) que posuem ortólogos em outras espécies detectados utilizando a ferramenta *Transmap*, obtidos a partir do UCSC Genome Browser. Os lncRNAs com ortólogos estão representados nas colunas e as espécies em que cada *transmap* foi encontrado nas linhas roxas. Para investigar a significância estatística, a proporção de lncRNAs que co-localizam com regiões genômicas associadas a ortólogos identificados por *transmaps* foi comparada às proporções resultantes do mapeamento de sequencias aleatórias. A significância foi dada por um valor de corte de *p-value* $\leq$ 0.05 no teste de Fisher (*).

4.3 Busca por elementos regulatórios associados às regiões promotoras (H3K4me3 e CpG) e evidências de 5'-CAP nos transcritos expressos em tecidos pancreáticos.

Devido à falta de informação sobre a biogênese desses RNAs não codificadores originados de regiões intrônicas e intergênicas, foram realizadas análises para a busca de elementos regulatórios no genoma que poderiam estar associadas à transcrição. Primeiro, foi visto a presença da trimetilação da lisina 4 da histona 3 (H3K4me3), uma modificação da cromatina descrita na literatura, associada com início de transcrição (Mikkelsen et al. 2007), na vizinhança das transcritos intrônicos e intergênicos. As coordenadas genômicas das marcas de cromatina H3K4me3 foram medidas em 13 linhagens celulares (Mikkelsen et al. 2007) e obtidas no UCSC Genome Browser. Apenas as marcas H3K4me3 que possuíam p-valor $< 10^{-5}$ foram utilizadas para limitar o ruído experimental. A marca H3K4me3 mais próxima relativa aos transcritos intrônicos e intergênicos foi selecionada e sua distância anotada. Como um controle, a mesma análise foi realizada usando 100 grupos randômicos compostos por sequências muito similares às expressas no tecido pancreático com o mesmo tamanho e mesmo conteúdo de CG. As coordenadas referentes aos RNAs codificadores de proteína expressos nos tecidos pancreáticos também foram analisadas. Para isso, as coordenadas das ESTs que representavam transcritos (intrônicos, intergênicos ou mRNAs) foram cruzadas com coordenadas correspondentes às marcas de H3K4me3 que estava presente em pelo menos uma das 13 linhagens cujas as marcas foram vistas. Após esta etapa a marca mais próxima foi anotada.

Como esperado, foi observada uma associação estatisticamente significante entre a localização de marcas H3K4me3 e mRNAs codificadores de proteínas, quando comparados com as 100 grupos controles (teste Kolmogorov-Smirnov, maior p-valor < 0.01). Resultados similares foram obtidos quando o grupo de transcritos intrônicos, mostrando uma associação

estatisticamente significativa com marcas associadas à região promotora H3K4me3 quando comparado com o grupo controle (Kolmogorov-Smirnov test, maior p-valor < 0.05) (Figura 18A). Quando os grupos de RNAs intrônicos e mRNAs foram comparados nenhuma diferença estatisticamente significante foi encontrada entre as duas distribuições, indicando que ambas as distribuições são similares. A mesma análise foi realizada para as sequências que mapeavam em regiões intergênicas. Apesar de ter sido observada uma frequência alta de marcas de cromatina perto das pontas conhecidas das sequências intergênicas Figura 18B), nenhuma diferença estatisticamente significante na distribuição foi encontrada quando comparada ao grupo randômico. Entretanto, nenhuma diferença na distribuição de marcas de cromatina H3K4me3 foi encontrada quando as sequências intergênicas foram comparadas com mRNAs, sugerindo que a distribuição dessas são similares. A análise de distribuição para ilhas CpG em relação às sequências intrônicas, intergênicas e mRNAs também foi realizada. A mesma estratégia de análise foi aplicada, as coordenadas genômicas de ilhas CpG (Gardiner-Garden and Frommer 1987) foram obtidas no UCSC Genome Browser, apenas ilhas CpG que apresentavam tamanho maior que 200bp e razão entre o número de dinucleotídeos CG observado em relação ao número esperado com base no números de C e G na sequência maior que 0.6 foram consideradas na análise. O único grupo que apresentou uma associação com ilhas CpG foi o grupo composto por sequências de mRNAs (Kolmogorov-Smirnov test, p-valor < 0.05) , tanto o grupo de RNAs intrônicos (Figura 18C) e RNAs intergênicos (Figura 18D) não apresentaram nenhuma associação com ilhas CpGs (Kolmogorov-Smirnov test, p-valor > 0.05) quando comparados ao grupo randômico. Quando com os grupos de sequências intrônicas e intergênicas foram comparados com o grupo de mRNAs, ambas distribuições mostraram-se diferentes daquela apresentada pelo grupo de mRNA (Kolmogorov-Smirnov test, p-valor < 0.001), sugerindo que a associação da distribuição das distâncias de ilhas CpG dos mRNAs são diferentes das encontradas para

ncRNAs intrônicos e intergênicos e que, talvez, a regulação por ilhas CpGs não seja o principal método de regulação apresentado no conjunto de sequências não codificadoras (intrônica e intergênica) estudadas.

A mesma análise foi realizada com dados de CAGE tags gerados de bibliotecas de RNA polyA(+) de 6 linhagens diferentes (RIKEN). Nenhuma dessas linhagens representava tecido pancreático. Para filtrar a grande quantidade de dados de CAGE *tags*, foram montados *clusters* e somente *clusters* que eram representados por 5 *tags* foram considerados para as análises. Após esta etapa a distância mais próxima dos CAGE *tags cluster* foi calculada em relação aos transcritos intrônicos, intergênicos, codificadores de proteína e as sequências provenientes dos grupos randômicos. Foi observado que ocorreu um enriquecimento significativo (KS teste, p-valor < 0.05) de CAGE *tags* a 1 kb de distância em relação ao começo putativo dos transcritos não codificadores intrônicos expressos em tecidos pancreáticos (Figura 18E). É importante deixar claro que todas essas análises foram realizadas tendo como coordenadas genômicas as pontas dos transcritos intrônicos e intergênicos, pois ainda não foram totalmente validados quanto ao seu tamanho.

Embora tenha sido observadad uma alta frequência de CAGE *tags* próxima às pontas de transcritos intergênicos, o enriquecimento não foi estatisticamente significativo quando comparado aos 100 conjuntos de sequências randômicas de mesmo conteúdo GC e tamanho (Figura 18F).

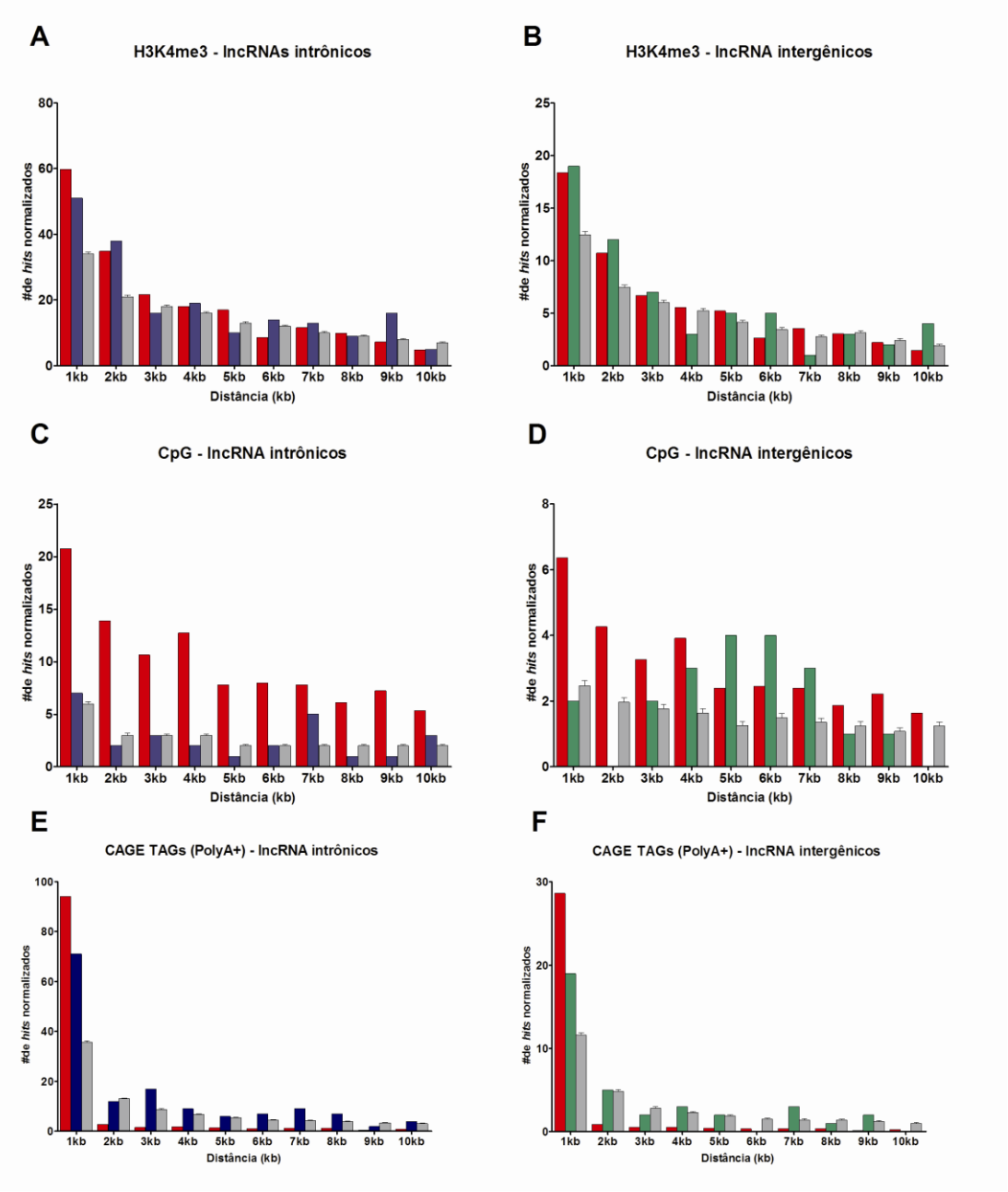

Figura 18: *Loci* gênicos de transcritos intrônicos estão enriquecidos com marcas de histonas associadas às regiões promotoras e transcritos com 5'CAP. A distribuição das distâncias (eixo X) de marcas de cromatina H3K4me3 associadas às regiões promotoras (painéis A e B), ilhas CpG (painéis C e D) e CAGE-*tags* (painéis E e F) relativa às coordenadas genômicas de lncRNAs intrônicos (barras azuis) e intergênicos (barras verdes) expressos em tecidos pancreáticos. Para tornar os dados comparáveis, o número de sequências codificadoras (barras vermelhas) e randômicas (barras cinzas) foi normalizado em relação ao conjunto total de sequências intrônicas (n=240) ou intergênicas (n=73). As barras cinza representam o número médio dos 100 grupos randômicos analisados e as barras de erro representam o erro padrão.

4.4 Identificação de assinaturas de expressão gênica de mRNAs e ncRNAs (intrônicos e intergênicos) correlacionadas com adenocarcinoma pancreático.

Para obter pistas de quais seriam as relevâncias biológica desses ncRNAs mapeados em regiões intrônicas e intergênicas no adenocarcinoma de pâncreas, nós investigamos quais seriam a expressão desses transcritos em amostras tumorais e não tumorais de tecidos pancreáticos de pacientes. Entretanto, a identificação de genes desregulados em células tumorais de pâncreas está, frequentemente, associada à presença de células estromais e inflamatórias (Gress et al. 1997; Logsdon et al. 2003). Uma reação desmoplásica parecida é observada na patologia de pancreatite crônica (Badea et al. 2008; Masamune and Shimosegawa 2009). Para favorecer a identificação de genes especificamente alterados no adenocarcinoma pancreático ductal, foi realizada uma análise comparando o perfil de expressão de 15 amostras de PDAC primários com 9 amostras histologicamente não tumorais combinados com oito amostras de pancreatite crônica. Com esta abordagem foram encontrados 147 transcritos diferencialmente expressos em adenocarcinoma pancreático comparados com tecidos não tumorais (FDR $\leq$ 10%). Esta assinatura de expressão era composta por 104 transcritos codificadores de proteínas e 43 lncRNAs, sendo 37 provenientes de regiões intrônicas e 9 de regiões intergênicas (Apêndice 2).

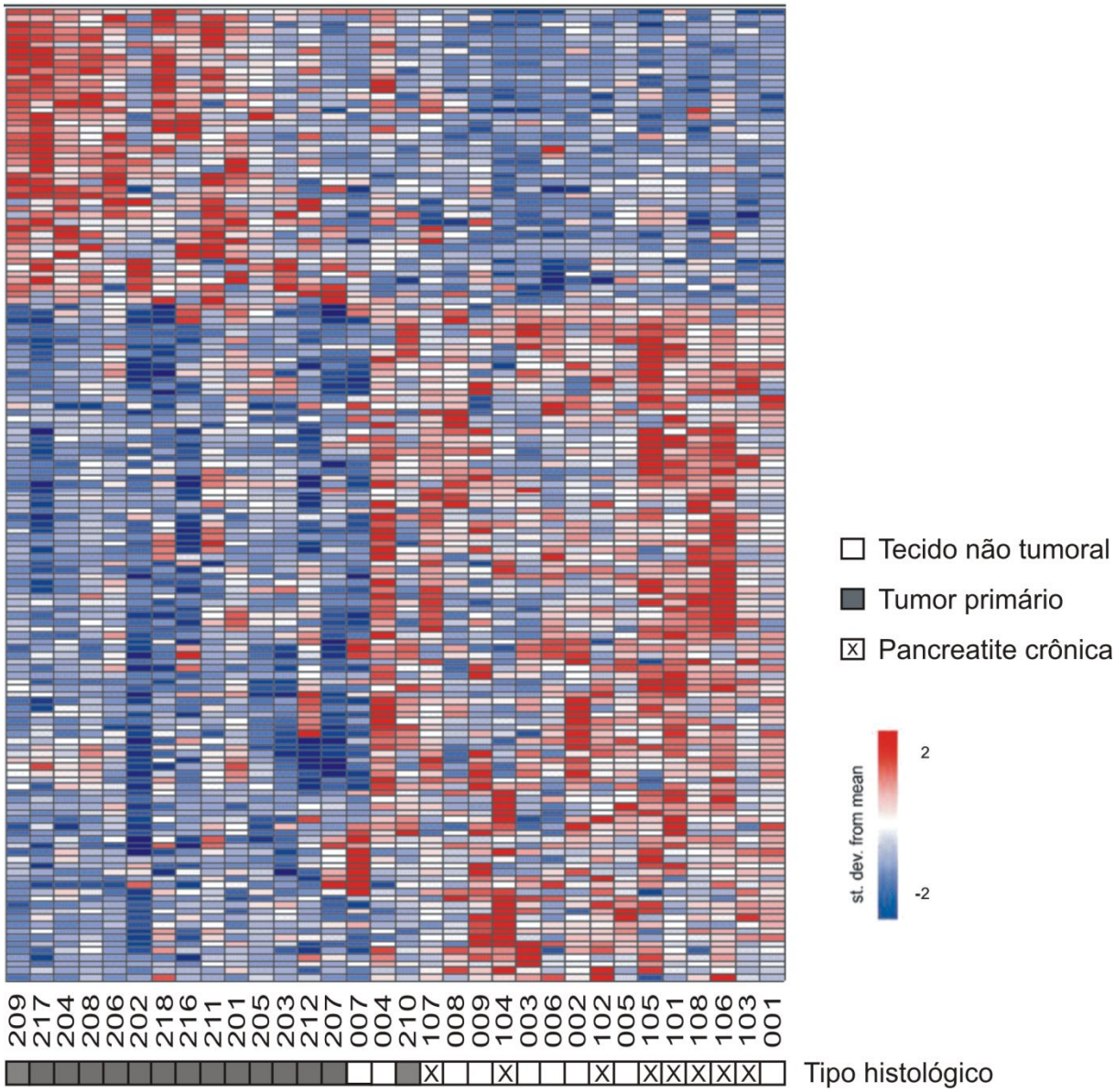

Figura 19: Assinatura de expressão gênica de PDAC. A análise estatística SAM *two-class* identificou 147 transcritos (linhas) diferencialmente expressos (FDR ≤ 10%) entre amostras de adenocarcinoma pancreática (quadrado cinza) e amostras de tecidos não tumorais borda adjacente ao tumor histologicamente não tumoral (quadrados brancos) e amostras de pancreatite crônica (quadrado marcado X) combinados. Cada coluna representa uma amostra diferente e os números estão relacionados com a identificação de cada paciente. Cerca de 29% (43/147) da assinatura é composta por transcritos não codificadores, sendo 9 mapeando em regiões intergênicas e 34 de regiões intrônicas. O nível de expressão de cada gene está representado pelo número de desvios-padrões acima (vermelho) ou abaixo (azul) do valor médio de expressão do transcrito entre todas as amostras. As amostras estão ordenadas de acordo com a correlação individual do perfil de cada amostra em relação ao perfil médio das amostras de tumor primário.

A Figura 19 mostra o cluster hierárquico das amostras tumorais e não tumorais. É possível observar que a assinatura consegue discriminar bem entre as amostras tumorais de adenocarcinoma primário e as amostras histologicamente não tumorais adjacentes ao tumor e pancreatite crônica. Somente uma amostra tumoral (210) foi ordenada juntamente com as amostras não tumorais. Provavelmente, a prevalência de componentes inflamatórios da amostra 210 poderia explicar o fato desta ter um perfil de expressão parecido com células de pancreatite crônica.

Devido ao fato do perfil de expressão apresentar tanto transcritos codificadores como não codificadores foi feita uma busca *in silico* a fim de comparar a lista de genes codificadores identificados como diferencialmente expressos com dados disponíveis na literatura, utilizando o *Pancreatic Expression Database* (Chelala et al. 2007). Dos 104 transcritos diferencialmente expressos identificados na assinatura de expressão, vinte e quatro transcritos (24/104, i.e. 23%) já haviam sido reportados em pelo menos um dos estudos que estavam no banco de dados utilizados. Desses 24 transcritos, dezessete apresentaram alterações nos níveis de expressão que estavam de acordo com outros estudos, três apresentaram uma concordância parcial na alteração do nível de expressão e 4 apresentaram um padrão de expressão invertido por outros estudos. A Tabela 5 mostra os transcritos diferencialmente expressos que foram encontrados em outros estudos de análise de expressão gênica depositados no banco de dados.

Tabela 5: Tabela de análise *in silico* comparando o conjunto de transcritos codificadores de proteína com aqueles identificados em outros trabalhos.

Gene Symbol	Entrez Gene ID	# Concordant studies/# Overlaping studies	Log2 (Fold Change # / #)															
			This study	a	b		c	d		e	f	g	h		i	j	k	l
			PDAC/NT +CP	PDAC/N	PDAC/CP	PDAC/NP	PDAC/N	PDAC/CP	PDAC/NP	PDAC/N	PDAC/NP	CL/NP	PDAC/CP	PDAC/NP	PDAC/NP	PDAC/N	PDAC/NP	PDAC/NP
S100P	6286	11/11	2.6	3.8	-	3	-	2.9	3.4	1.6	3.1	4.1	4.1	4.5	-	>2.3	7.3	-
COL1A1	1277	4/4	1.34	-	-	2.32	-	0.16	2	-	-	-	-	-	-	>2.32	-	-
ARPC1B	10095	2/2	0.96	-	-	-	-	0.93	1.84	-	-	-	-	-	-	-	-	-
SDC1	6382	2/2	0.71	-	-	-	-	1.04	1.89	-	-	-	-	-	-	-	-	-
TCTA	6988	2/2	-0.54	-	-	-	-	-1.36	-2.18	-	-	-	-	-	-	-	-	-
S100A6	6277	2/2	1.2	1.42	-	2.81	-	-	-	-	-	-	-	-	-	-	-	-
TGFBI	7045	2/3	1.01	-	-	-	-	-0.1	2.39	-	-	-	-	-	2.32	-	-	-
DLC1	10395	1/1	-0.7	-	-	-	-	-	-	-	-	-	-	-	-	<-2.32	-	-
TLE1	7088	1/1	-0.9	-	-	-	-	-	-	-	-	-	-	-	-	<-2.32	-	-
ACTG1	71	1/1	0.4	1.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AXIN2	8313	1/1	-0.4	-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ITGB5	3693	1/1	0.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.1
MTA1	9112	1/1	-0.6	-	-	-	-	-	-	-	-	-	-	-	-	<-2.32	-	-
MBOAT2	129642	1/1	0.7	-	-	-	-	-	-	-	-	-	-	-	-	>2.32	-	-

TIMP1	7076	1/1	2.62	-	-	-	UP	-	-	-	-	-	-	-	-	-	-	-
TMEM141	85014	1/1	1.09	1.28	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TUSC3	7991	1/1	0.91	1.51	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ZSCAN18	65982	1/1	-1.14	-	-	-	-	-	-	-	-	-	-	-	-	<-2.32	-	-
CD14	929	1/2	-0.41	-	-	-	-	-0.14	2.65	-	-	-	-	-	-	-	-	-
BCAR3	8412	0/1	-0.51	1.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NOTCH4	4855	0/1	-0.89	-	-	-	UP	-	-	-	-	-	-	-	-	-	-	-
TRA1	7184	0/1	1.05	-	-	-	-	-	-	-	-	-	-	-	-1.58	-	-	-
HPN	3249	0/2	1.25	-	-	-	-	-3.47	-4.32	-	-	-	-	-	-	-	-	-
STX6	10228	0/2	1.37	-	-9.35	-5.64	-	-	-	-	-	-	-	-	-	-	-	-

PDAC: Adenocarcinoma pancreático ductal, NT: tecido não tumoral adjacent ao tumor, NP : Tecido Normal (*bulk tissue*), Normal: Células ductais normais (tecido microdissecado), CP : Pancreatite Crônica, CL : Linhagens celulares neoplásicas.

[a]Buchholz M *et al*, 2005 (Buchholz et al. 2005); [b]Crnogorac-Jurcevic T *et al*, 2005 (Crnogorac-Jurcevic et al. 2005); [c]Crnogorac-Jurcevic T *et al*, 2003 (Crnogorac-Jurcevic et al. 2003); [d]Crnogorac-Jurcevic T *et al*, 2002 (Crnogorac-Jurcevic et al. 2002); [e]Friess H *et al*,2003 (Friess et al. 2003); [f]Grutzmann R *et al*, 2004 (Grutzmann et al. 2004); [g]Iacobuzio-Donahue CA *et al*, 2002 (Iacobuzio-Donahue et al. 2002); [h]Logsdon CD *et al*, 2003 (Logsdon et al. 2003), [i]Lu *et al*, 2004 (Lu et al. 2004); [j]Nakamura T *et al*, 2004 (Nakamura et al. 2004); [k]Segara D *et al*, 2005 (Segara et al. 2005) ; [l]Tan ZJ *et al*,2003 (Tan et al. 2003)

Alguns dos transcritos codificadores apresentados na Tabela 5 e na assinatura associada ao adenocarcinoma pancreático ductal já foram estudados e propostos como biomarcadores como *S100A6*, *TIMPI*, *NF-κB*, *VCL* e *S100P* (Crnogorac-Jurcevic et al. 2003; Iacobuzio-Donahue et al. 2003; Shekouh et al. 2003; Buchholz et al. 2005; Ohuchida et al. 2007; Deng et al. 2008; Harsha et al. 2009). Ainda é importante enfatizar que o transcrito codificador *S100P* foi encontrado como aumentado em 11 estudos de expressão gênica diferentes e ainda, recentemente, foi avaliado como um bom marcador para diagnóstico de PDAC em amostras de FNA (*Fine-neelde aspiration*) sendo detectado em mais de 90% das amostras tumorais (n=40) (Liu et al. 2012). Outro gene interessante identificado foi o *MBOAT*, e, o aumento da expressão deste transcrito em amostras de adenocarcinoma pancreático está inversamente correlacionado com o tempo de sobrevida do paciente em câncer pancreático (Badea et al. 2008).

Para investigar quais seriam as vias que estavam mais alteradas na assinatura associada ao adenocarcinoma pancreático foi realizada uma análise de enriquecimento de vias utilizando a ferramenta DAVID (*Database for Annotation, Visualization and Integrated Discovery*) (Hosack et al. 2003). Esta análise teve como objetivo identificar as vias alteradas tanto com transcritos codificadores de proteínas ou transcritos mapeados em regiões intrônicas. Para isso os lncRNAs intrônicos foram anotados de acordo com o *locus* gênico de onde estes mapeiam. Somente as vias alteradas que apresentaram EASE *score* < 0.05 foram consideradas como enriquecidas na análise (Hosack et al. 2003).

Dos 104 codificadores de proteínas diferencialmente expressos em amostras de tumor pancreático e foram encontradas as vias relacionadas que contem proteínas associadas a 17 GO (*Gene Ontology*) e uma via da base de dados do KEGG *pathway* (Apêndice 3). Para retirar a redundância apresentada pelos termos de *Gene Ontology* (GO), foi utilizado o termo de GO de maior hierarquia como mostrado na Figura 20 que foi montada no site AmiGO

(http://amigo.geneontology.org/cgi-bin/amigo/visualize?mode=client&session_id=) a partir dos termos encontrados como enriquecidos na análise pelo DAVID. Na Tabela 6 estão representados os principais GOs da lista de análise de enriquecimento e vias encontradas em outras bases de dados, a lista completa está no Apêndice 3.

Tabela 6: Principais categorias representadas da análise de enriquecimento (DAVID) de genes diferencialmente expressos na assinatura de amostras de tumores primários comparados à pancreatite crônica e amostras não tumorais.

Categoria	**Termos**	p-valor*	**Genes**
Genes codificadores de proteína			
GOTERM_BP_FAT GO:0019048	Virus-Host Interaction	0,005	THOC7, THOC2, INSR, THOC3
GOTERM_BP_FAT GO:0050658	RNA transport	0,007	THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0006403	RNA Localization	0,010	THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0015931	Nucleobase, containing compound Transport	0,010	THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_CC_FAT GO:0070161	Anchoring Junction	0,044	TRIP6, PERP, TRIM25, VCL, SDC1, ARPC2, DLC1, CDH1
GOTERM_CC_FAT GO:0000346	Transcription Export Complex	0,025	THOC7, THOC2, THOC3
GOTERM_CC_FAT GO:0016323	Basolateral Plasma Membrane	0,039	TRIP6, NOTCH4, TRIM25, VCL, SDC1, ARPC2, DLC1, CDH1

*corrigido por **EASE SCORE**

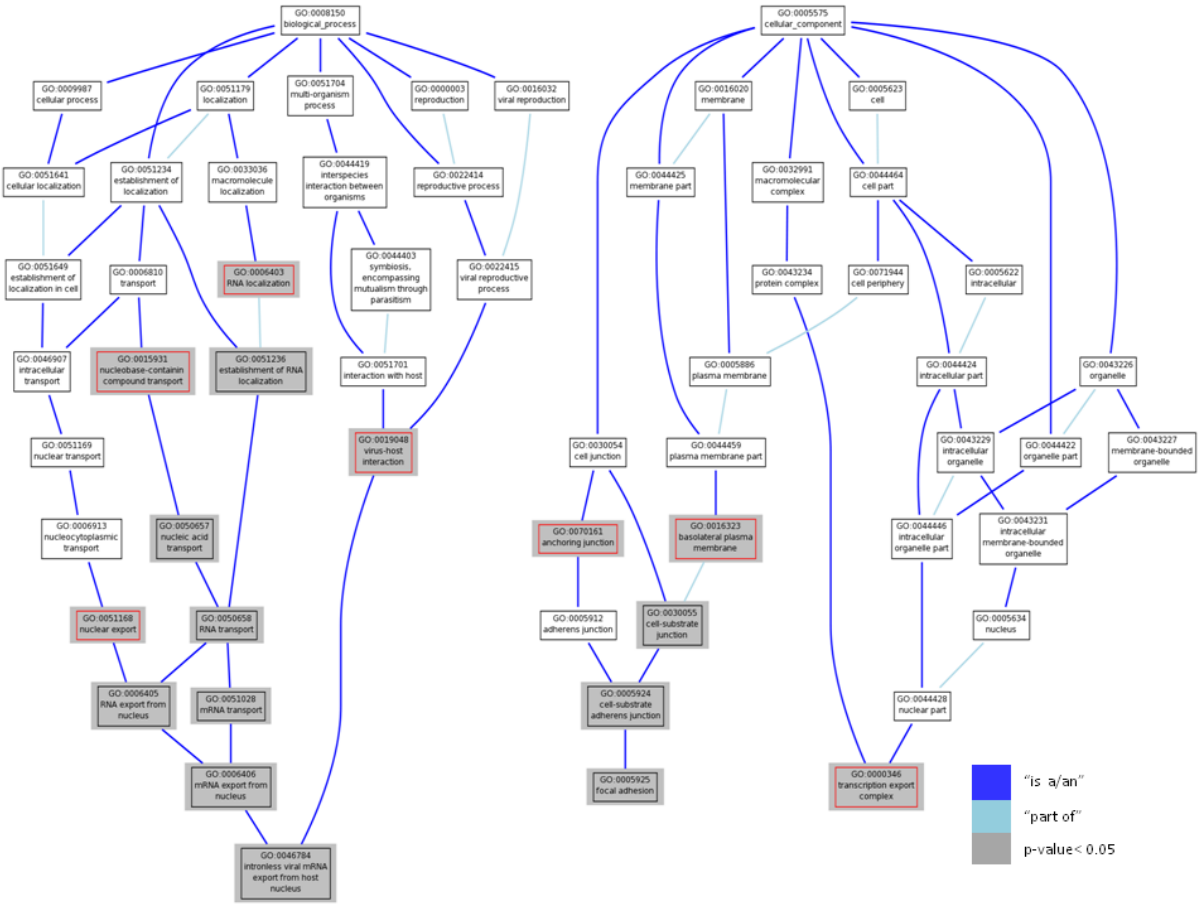

Figura 20: Visualização da estrutura de termos de GO associados ao adenocarcinoma primário de pâncreas. As categorias estão apresentadas em nódulos (quadrados) e as relações entre os nódulos estão representados por conexões (linhas azuis). A relação delineada pela linha azul escura ("*is a/an*") indica que o nódulo é um subtipo do nódulo acima. A linha azul-clara ("*is part of*"") indica que o nódulo é parte do nódulo acima. Termos mais gerais estão representados na parte de cima do gráfico e GOs mais específicos estão apresentados na parte mais baixa do gráfico. Os nódulos representados em cinza são os termos de GO que foram encontrados como enriquecidos na análise realizada pelo DAVID. Termos marcados em vermelho representam termos mais gerais que estão especificados na Tabela 6.

A desregulação de genes codificadores de proteínas envolvidos na via adesão focal (Figura 20) já havia sido reportada em outros trabalhos na literatura (Badea et al. 2008). Dentre as categorias mostradas na Tabela 6, a classe de localização de RNA aparece como

enriquecida e é composta por genes associados ao *TREX-complex* (*THOC2* e *THOC3*). Aumento da expressão de genes associados a esse complexo (Thoc1) já foi descrita em câncer de mama e correlaciona com o tamanho e progressão tumoral (Guo et al. 2005). Ainda o aumento da expressão de *THOC2* já foi reportada em amostras PanIN-3 (*Pancreatic intraepithelial neoplasias*) em relação à amostras ductais pancreáticas normais (Buchholz et al. 2005). Todos esses dados sugerem que a modulação de *TREX-complex* poderia estar associada ao adenocarcinoma pancreático.

A alteração de alguns genes associados às categorias representadas na Tabela 6 já foram associados com outras lesões pré-malignas associadas ao adenocarcinoma pancreático como *PanIN* (*Pancreatic Intraepithelial Neoplasias*) e *IPMN* (*Intraductal Papillary Mucinous Neoplasms*) como pode ser vista na Figura 21. Interessantemente, o transcrito codificador VCL foi detectado como aumentado em pacientes com adenocarcinoma pancreático no perfil de expressão deste estudo também foi identificado como aumentado em lesões pré-malignas PanIN-1B/2 quando comparados com tecidos normais (Prasad et al. 2005). Ainda, a proteína foi detectada em amostras não tumorais pancreáticas (Hu et al. 2004) e na urina de indivíduos saudáveis (Adachi et al. 2006), podendo se tornar um possível alvo para estudos de marcadores moleculares no adenocarcinoma pancreático.

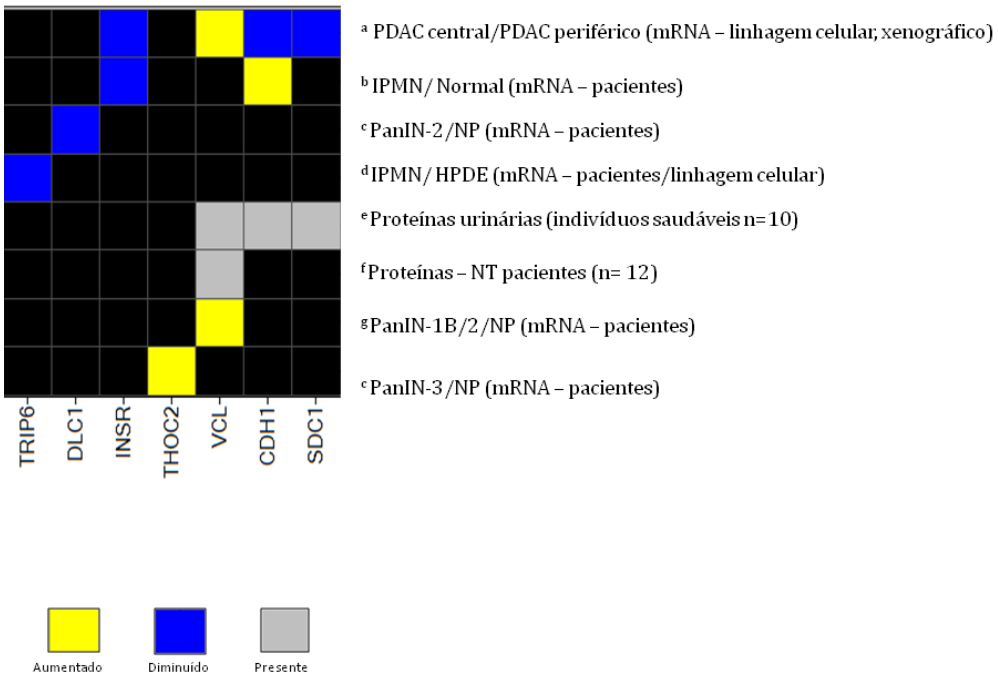

Figura 21: Genes envolvidos com categorias enriquecidas também estão associados a outros tipos de lesões que antecedem PDAC. Dos 15 genes representados na Tabela 6, sete (colunas) já foram associados com lesão pré-malignas. As comparações estão descritas nas linhas. Genes aumentados estão representados em amarelo e genes diminuídos em azul. NT: tecido não tumoral adjacent ao tumor, PDAC: Adenocarcinoma pancreático ductal, NP : Tecido Normal (*bulk tissue*), Normal: Células ductais normais (tecido microdissecado), PanIN (*Pancreatic Intraepithelial Neoplasias*) e IPMN (*Intraductal Papillary Mucinous Neoplasms*). a (Nakamura et al. 2007), b (Sato et al. 2004), c (Buchholz et al. 2005), d (Jones et al. 2008), e (Consortium 1998; Adachi et al. 2006), f (Adachi et al. 2006), g (Consortium 1998; Prasad et al. 2005).

A análise de enriquecimento de categorias foi realizada para os *loci* gênicos correspondentes aos transcritos intrônicos, entretanto nenhuma categoria enriquecida foi encontrada.

O IPA (*Ingenuity Pathway Analysis*) foi utilizado para identificar vias e redes regulatórias de genes representadas no conjunto de mRNAs codificadores de proteínas que compunham a assinatura associada ao PDAC. Interessantemente, a via mais enriquecida denominada "movimento celular, sinalização de interação célula-célula e sistema endócrino" (p-valor < 10^{-43}) foi composta por 23 transcritos diferencialmente expressos e compunha genes representados nas categorias enriquecidas usando o DAVID (*ARPC2*, *CDH1*, *NOTCH4*

e *VCL*) e na análise *in silico* de genes diferencialmente expressos na Tabela 5 (*ARPC1B*, *ACTG1*, *COL1A1*, *ITGB5*, *MTA1*, *NOTCH4* e *TIMP1*) sugerindo que a desregulação desta via está associada ao PDAC. Outras redes de interação gênica que foram encontradas foram "movimento celular, função e desenvolvimento do sistema esquelético e muscular e resposta inflamatória" (17 genes, p-valor < 10^{-29}) e "metabolismo de carboidratos, bioquímica de pequenas moléculas e doenças infecciosas" (16 genes, p-valor < 10^{-28}) também foram identificadas.

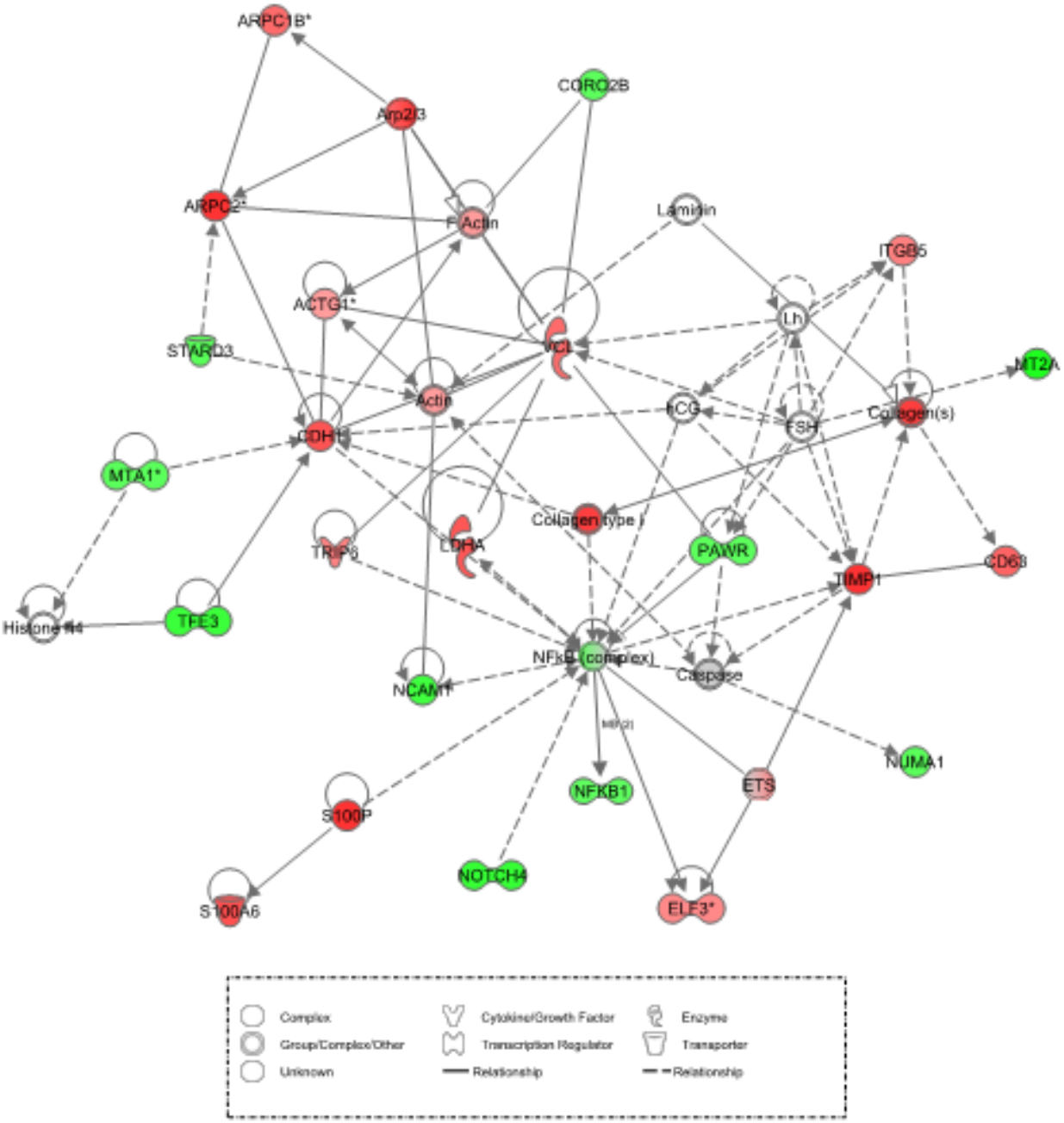

Figura 22: Genes modulados no adenocarcinoma pancreáticos estão envolvidos em movimento celular, interação célula-célula e sistema endócrino. *Ingenuity Pathway Analysis* foi utilizado para identificar redes de interação gênica enriquecidas ao longo do conjunto de 104 codificadores de proteínas diferencialmente expressos em amostras de adenocarcinoma pancreático. A rede mais enriquecida (*p-value* < 10^{-43}) é composta por 23 genes diferencialmente expressos medidos pelo experimento de microarranjo. Vermelho indica alta expressão e verde, baixa expressão em tecidos tumorais em relação à pancreatite crônica e tecidos não tumorais pancreáticos.

4.5 Identificação de genes correlacionados com a metástase do câncer de pâncreas.

Uma das características do câncer pancreático é a alta prevalência da doença metastática, a qual as bases moleculares ainda são pouco conhecidas. Para identificar transcritos codificadores de proteínas e RNAs não codificadores longos com o nível de expressão correlacionado com o fenótipo pancreático, foram comparados o perfil de expressão de 15 amostras de pacientes com adenocarcinoma pancreático e seis amostras de metástases distantes originados de adenocarcinoma pancreático primário. As amostras pancreáticas foram coletadas em sítios diferentes ao pâncreas (1 peritôneo, 1 gânglio e 4 de fígado) de diferentes pacientes. Utilizando o SAM com FDR ≤ 5%, foi identificada uma assinatura do perfil de expressão associada à metástase contendo 355 transcritos diferencialmente expressos (Figura 23). Destes, 221 são codificadores de proteína e 134 correspondem á transcritos não codificadores (134/355, 38% da assinatura), dos quais 101 mapeavam em regiões intrônicas, 27 intergênicas e 6 a lncRNAs conhecidos. Interessantemente, cento e onze foram identificados como alterados em outros trabalhos que estudavam doenças pancreáticas, como pancreatite crônica, PDAC primário ou lesões pré-malignas, que estão mostrados na Tabela 7. Esta análise foi realizada utilizando a base de dados de expressão no site http://www.pancreasexpression.org/ (Cutts et al. 2010).

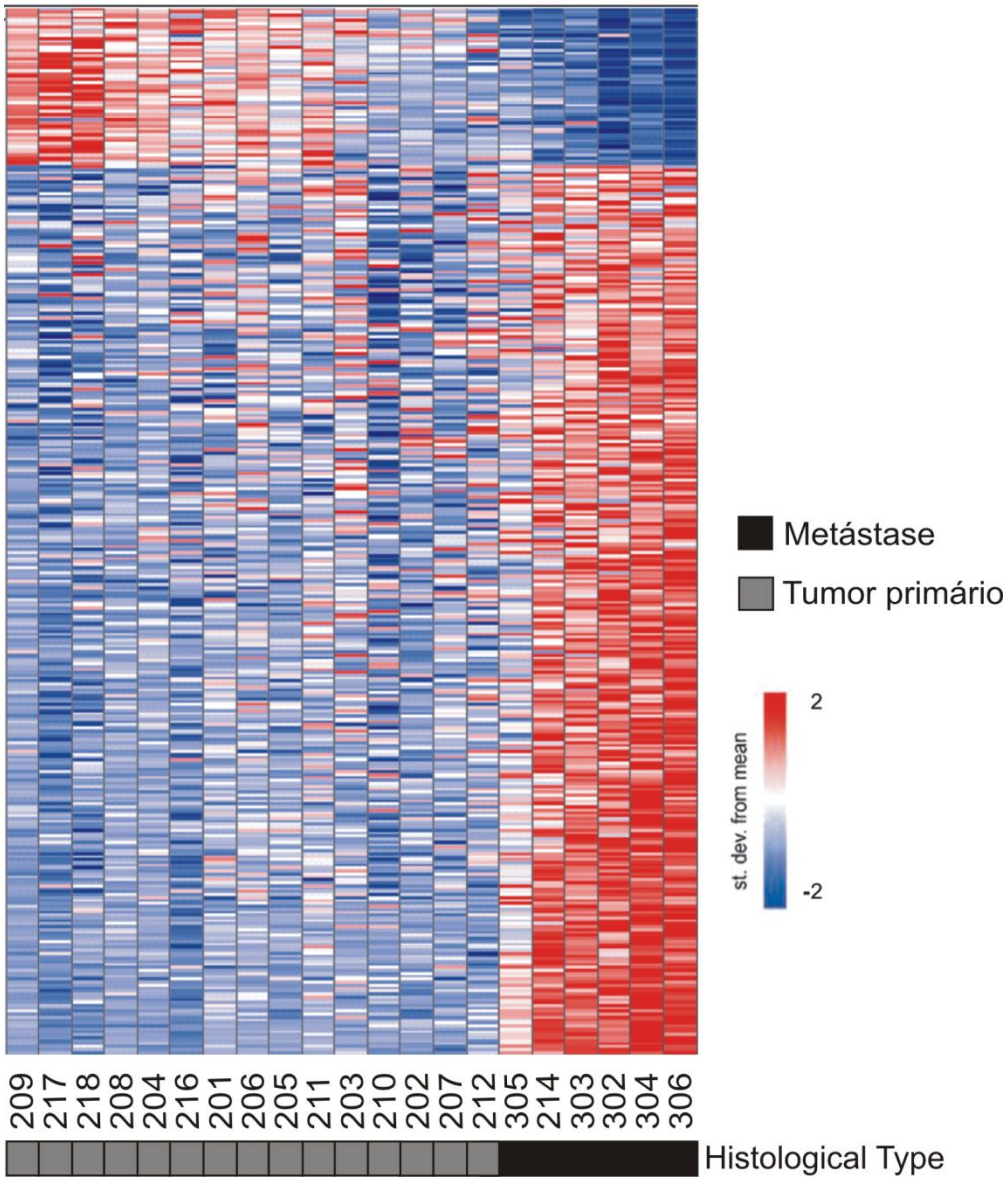

Figura 23: Assinatura de expressão gênica associada à metástase de PDAC. A análise estatística SAM *two-class* identificou 355 transcritos (linhas) diferencialmente expressos (FDR ≤ 5%) entre amostras de adenocarcinoma pancreática (quadrado cinza) e amostras de tecidos metastáticos (quadrado preto). Cada coluna representa uma amostra diferente e os números estão relacionados com a identificação de cada paciente. Cerca de 38% (134/355) da assinatura é composta por transcritos não codificadores, sendo 27 mapeando em regiões intergênicas e 101 de regiões intrônicas. O nível de expressão de cada gene está representado pelo número de desvios-padrões acima (vermelho) ou abaixo (azul) do valor médio de expressão do transcrito entre todas as amostras. As amostras estão ordenadas de acordo com a correlação individual do perfil de cada amostra em relação ao perfil médio das amostras de tumor primário.

Tabela 7: Comparação dos codificadores identificados como alterados em metástases pancreáticas com outros trabalhos na literatura.

Comparação	Nº de genes	Genes	Trabalhos
Pancreatite crônica *versus* Normal	22	S100A8, RAB4A, TYRO3, TCTA, THBS1, SRGN, B2M, RAN, RAB31, EMP1, ECE1, MAPK3, DDX11, PPWD1, EWSR1, DPYSL2, MYH11, CD14, LAP3, HIF1A, ARPC1B, FTL	(Friess et al. 2003; Shen et al. 2004; Crnogorac-Jurcevic et al. 2005; Chen et al. 2007)
Tratamento com drogas	33	RAB31, MXRA7, PICALM, IGFBP7, CBFB, SIVA1, TCF12, NAV3, PURB, MLH1, BCAR3, CHD2, GOSR2, DNAJC9, TPD52L2, RBM19, EFNB2, TFPI, OSBPL3, THBS1, SRGN, EMP1, ECE1, DPYSL2, AREG, ARRDC4, FRMD4B, S100A4, PRSS12, ERBB3, RAB2A, ZSCAN18, HIF1A	(Song et al. 2008; Arumugam et al. 2009; Selga et al. 2009; Wang et al. 2009)
Lesões pré-malignas (IPNM ou PanIN) *versus* Normal	28	LMO2, TDRD6, ZNF438, INSR, ARRDC4, PRSS12, IGFBP7, CBFB, SIVA1, BCAR3, TPD52L2, EMP1, DAP3, RPL15, LDHA, SMAD5, MAX, NUDT16L1, PTPRA, THOC2, POMP, IFI6, CTBP1, NCAM1, B2M, FTL, S100A4, PALLD	(Sato et al. 2004; Buchholz et al. 2005; Prasad et al. 2005; Pogue-Geile et al. 2006)
PDAC *versus* lesões benignas ou Pancreatite crônica	21	THBS1, RAB4A, TYRO3, TCTA, MYH11, CD14, RAB31, SRGN, ECE1, DPYSL2, S100A8, RAB3B, NMI, DDX11, PPWD1, EWSR1, ARPC1B, HIF1A, EMP1, B2M, FTL	(Friess et al. 2003; Shen et al. 2004; Yu et al. 2005; Chen et al. 2006; Pilarsky et al. 2008)
PDAC *versus* Normal	53	RAB4A, TYRO3, TCTA, ZNF438, PRSS12, RPS20, CCAR1, ITGB7, MAP3K7IP1, ZSCAN18, MAPK3, PHB, HSP90B1, NFATC2IP, SRGN, EFNB2, DDX3X, TBC1D30, TAF8, TRIM25, MFSD8, NOTCH4, BYSL, ANKRD40, SH3PXD2A, AREG, CBFB, BCAR3, PTPRA, IFI6, DDX11, PPWD1, EWSR1, ARPC1B, PALLD, MYH11, EMP1, RAB13, S100A4, ECE1, B2M, IGFBP7, LAP3, FTL, CD14, S100A8, HIF1A, PMEPA1, TK1, RAB2A, MMP2, THBS1, RAB31	(Crnogorac-Jurcevic et al. 2001; Crnogorac-Jurcevic et al. 2002; Iacobuzio-Donahue et al. 2002; Crnogorac-Jurcevic et al. 2003; Friess et al. 2003; Tan et al. 2003; Grutzmann et al. 2004; Lu et al. 2004; Nakamura et al. 2004; Shen et al. 2004; Buchholz et al. 2005; Chen et al. 2005;

			Crnogorac-Jurcevic et al. 2005; Pogue-Geile et al. 2006; Badea et al. 2008; Jones et al. 2008; Zhang et al. 2009; Zhang et al. 2009; Gnjatic et al. 2010)
PDAC central *versus* PDAC periférico	25	MMP2, TK1, PMEPA1, EMP1, PALLD, HSP90AA1, NCAM1, POMP, GPATCH4, RFC4, FERMT1, PHB, RAB31, S100A8, SRGN, NFATC2IP, TFPI, RPL15, ACAD11, BCL6, LEPROT, INSR, LMO2, RAB4A, RAB4A	(Nakamura et al. 2007)
PET *versus* Normal	17	HSP90B1, LMO2, JAK1, MAP4K5, FNDC3B, HSP90AA1, RAB2A, MAPK3, EIF2AK2, ARNT, CEP78, STARD3, SIVA1, MYH11, CD14, IGFBP7, EMP1	(Maitra et al. 2003; Capurso et al. 2006)
Identificação de proteínas no plasma, urina ou pâncreas	28	RAB2A, TFPI, GNB1, CHMP5, CD63, GNAQ, EIF4EBP2, CHMP2A, ITGAX, NUDT1, PHB, PACSIN2, DPYSL2, RAB3B, RAB13, S100A4, ECE1, B2M, MMP2, IGFBP7, HSP90AA1, LDHA, LAP3, FTL, THBS1, CD14, HSP90B1, S100A8	(Anderson et al. 2004; Gronborg et al. 2004; Hu et al. 2004; Adachi et al. 2006)

Informações extraídas do *Pancreatic Database Expression*.

Utilizando a mesma abordagem anterior, fomos verificar quais seriam as categorias enriquecidas que estavam representadas nos transcritos diferencialmente expressos. Os transcritos codificadores de proteína diferencialmente expressos em amostras metastáticas apresentaram 12 categorias de GO enriquecidas. Dessas, cinco foram escolhidas, pois apresentavam GO mais gerais e estão apresentadas na Tabela 8. As principais categorias estavam envolvidas com categorias de “transporte de ácido nucleico” e “localização de RNA” ($p < 0.03$; THOC7, THOC2, RAN, NUP85, THOC3 e NUP88). Uma análise similar realizada com *locus* gênico contendo lncRNAs intrônicos que compunham a assinatura mostrou enriquecimento em categorias pertencendo à sinalização de *MAPK* quinases e termos de GO relacionados ao *phosphate metabolic process*, *non-membrane-bounded organelle* e *actin-filament-based process* (Tabela 8). A fim de retirar a redundância entre os termos foi utilizado o termo de GO de maior hierarquia como mostrado na Figura 24 que foi montada no site

AmiGO (http://amigo.geneontology.org/cgi-bin/amigo/visualize?mode=client&session_id=) a partir dos termos encontrados como enriquecidos na análise pelo DAVID.

Tabela 8: Principais categorias representadas da análise de enriquecimento (DAVID) de genes diferencialmente expressos na assinatura de amostras de tumores primários comparados a metástases pancreáticas.

Termos	Categorias	p-valor*	Genes
codificador de proteína			
BIOCARTA h_hifPathway	Hypoxia-Inducible Factor in the Cardiovascular System	0.011	ARNT, LDHA, HSP90AA1, HIF1A, VHL
GOTERM_BP_FAT GO:0006403	RNA localization	0.026	NUP88, THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0015931	nucleobase, nucleoside, nucleotide and nucleic acid transport	0.026	NUP88, THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0051168	nuclear export	0.019	THOC7, STRADA, THOC2, RAN, THOC3
GOTERM_BP_FAT GO:0019048	virus-host interaction	0.038	THOC7, THOC2, INSR, THOC3
GOTERM_CC_FAT GO:0031982	Vesicle	0.009	RAB13, PRSS12, CAMKV, RAB2A, ECE1, RAN, WIPF1, TMEM33, PICALM, SRGN, TGOLN2, HSP90AA1, GGNBP2, MYH11, RAB4A, PACSIN2, RAB3B, HSP90B1, PIK3C2A, THBS1
***locus* gênico de transcritos intrônicos**			
GOTERM_CC_FAT GO:0043228	non-membrane-bounded organelle	0.016	ABI2, GPHN, ITPR1, SORBS1, TYK2, NDRG1, DST, MCPH1, USH1C, MAEA, BBS5, SLC4A7, RAPH1, CNN3, NR2C2, DMD, DAZAP1, MAP3K1, MAP2K5, PHF12, NOP5/NOP58, ATF3, ALMS1, STON2, DAPK1, MYO5A, MAPK1
KEGG_PATHWAY hsa04010	MAPK signaling pathway	0.021	ARRB1, MAP3K1, ATF2, MAP3K14, MAP2K5, TGFBR2, PPP3CB, RAPGEF2, MAPK1
GOTERM_BP_FAT GO:0006793	phosphorus metabolic process	0.042	ABI2, ENPP2, PTEN, CSNK1D, MAP3K1, MAP2K5, TYK2, PPP3CB, PPP2R2A, PASK, MAP3K14, TNK2, TGFBR2, DAPK1, MAPK1
GOTERM_BP_FAT GO:0030029	actin filament-based process	0.048	ABI2, SORBS1, PACSIN2, CNN3, MYO5A, DST

*corrigido por **EASE SCORE**

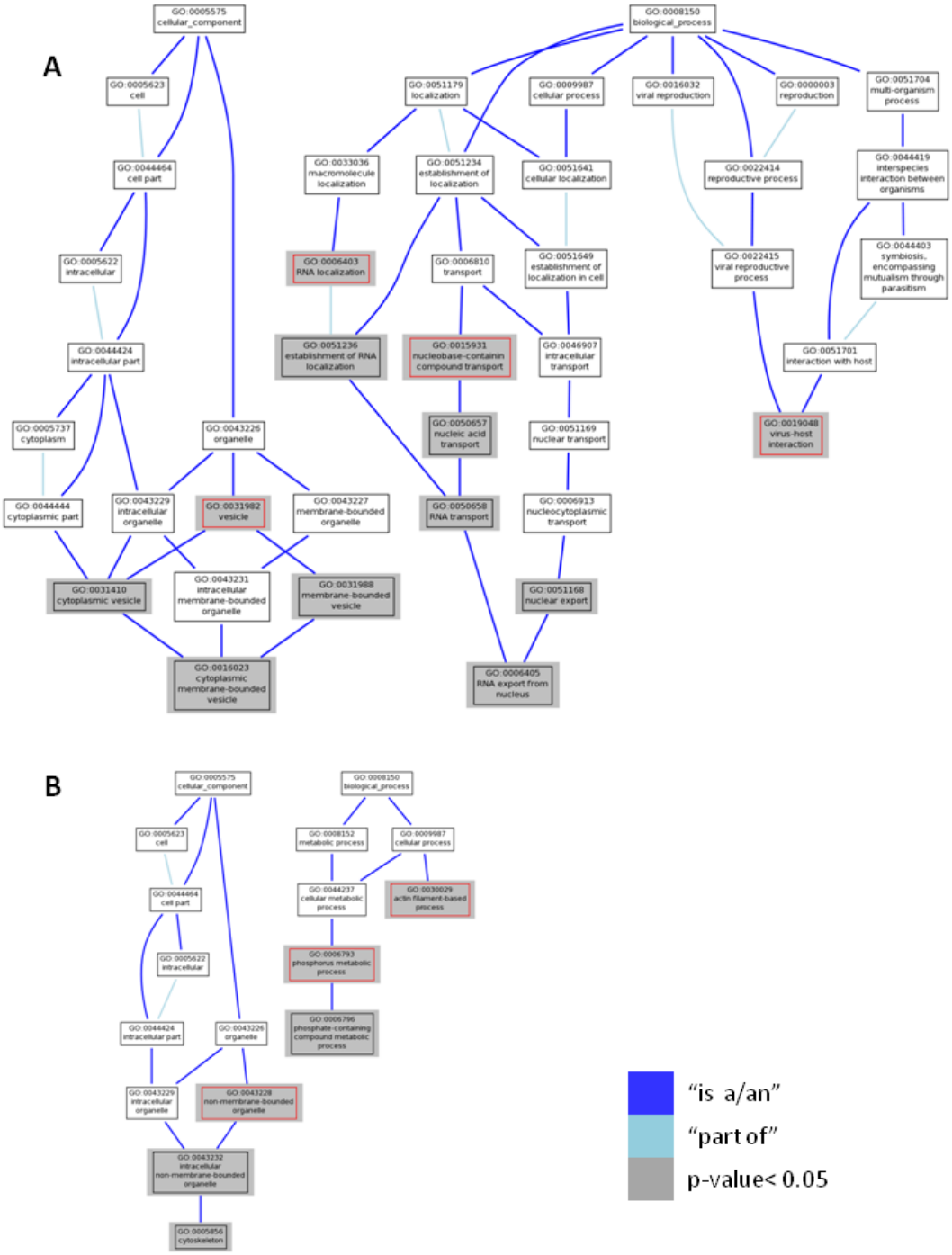

Figura 24: Visualização da estrutura de termos de GO associados ao adenocarcinoma metastático de pâncreas. Representação referente aos GOs encontrados como enriquecidos no subcojunto de **A-)codificadores de proteína** e **B-)** ***loci*** **gênicos de transcritos intrônicos**. As categorias estão apresentadas em nódulos (quadrados) e as relações entre os nódulos estão representados por conexões (linhas azuis). A relação delineada pela linha azul escura ("*is a/an*") indica que o nódulo é um subtipo do nódulo acima. A linha azul-clara ("*is part of*"") indica que o nódulo é parte do nódulo acima. Termos mais gerais estão representados na parte de cima do gráfico e GOs mais específicos estão apresentados na parte mais baixa do gráfico. Os nódulos representados em cinza são os termos de GO que foram encontrados como enriquecidos na análise realizada pelo DAVID. Termos marcados em vermelho representam termos mais gerais que estão especificados na Tabela 8.

Análises utilizando IPA também foram realizadas e identificaram redes gênicas enriquecidas para o conjunto de codificadores de proteínas diferencialmente expressos. A rede mais enriquecida de codificadores de proteínas mRNAs possuía 24 genes diferencialmente expressos (p-valor < 10^{-41}) incluindo genes relacionados a "movimento celular, expressão gênica e tráfico de células imunes" (Figura 25). Dentre estes, o aumento de expressão de genes *S100A4*, *NCAM1* e *LIMK1* já haviam sido associados com metástase em câncer de pâncreas (Tabata et al. 2009; Vlecken and Bagowski 2009; Aloysius et al. 2010). Os outros transcritos desta via ainda não haviam sido associados com a metástase pancreática, entretanto muitos deles já haviam sido descritos como envolvidos em outras malignidades e metástases em outros tipos de tumores Apêndice 6.

Outras redes de interação gênica foram encontradas para os codificadores de proteínas desregulados na metástase pancreática estavam relacionadas com "ciclo celular, doenças genéticas e metabólicas" (21 genes, $p < 10^{-33}$), "desenvolvimento de tecido, sistema cardiovascular e embrionário" (21 genes, $p < 10^{-30}$) e "câncer, morfologia tumoral e doença genética" (19 genes, $p < 10^{-29}$). Utilizando os *loci* gênicos dos lncRNAs intrônicos, a análise do IPA identificou 42 genes que estavam correlacionados com morte celular, sendo 19 com expressão diminuída e 23 com o nível de expressão aumentados em amostras de metástase pancreática. Isto é importante pelo fato da perturbação das vias envolvidas com morte celular já haviam sido associadas ao fenótipo metastático no adenocarcinoma pancreático (Jones et al. 2008; Ruckert et al. 2010). Interessantemente, foram encontrados 6 transcritos intrônicos que mapeavam em *loci* gênicos que estavam associados ao processo de apoptose (*ATF2*, *TGFβR2*, *MAP2K5*, *MAP3K1*, *DAPK1* e *PTEN*).

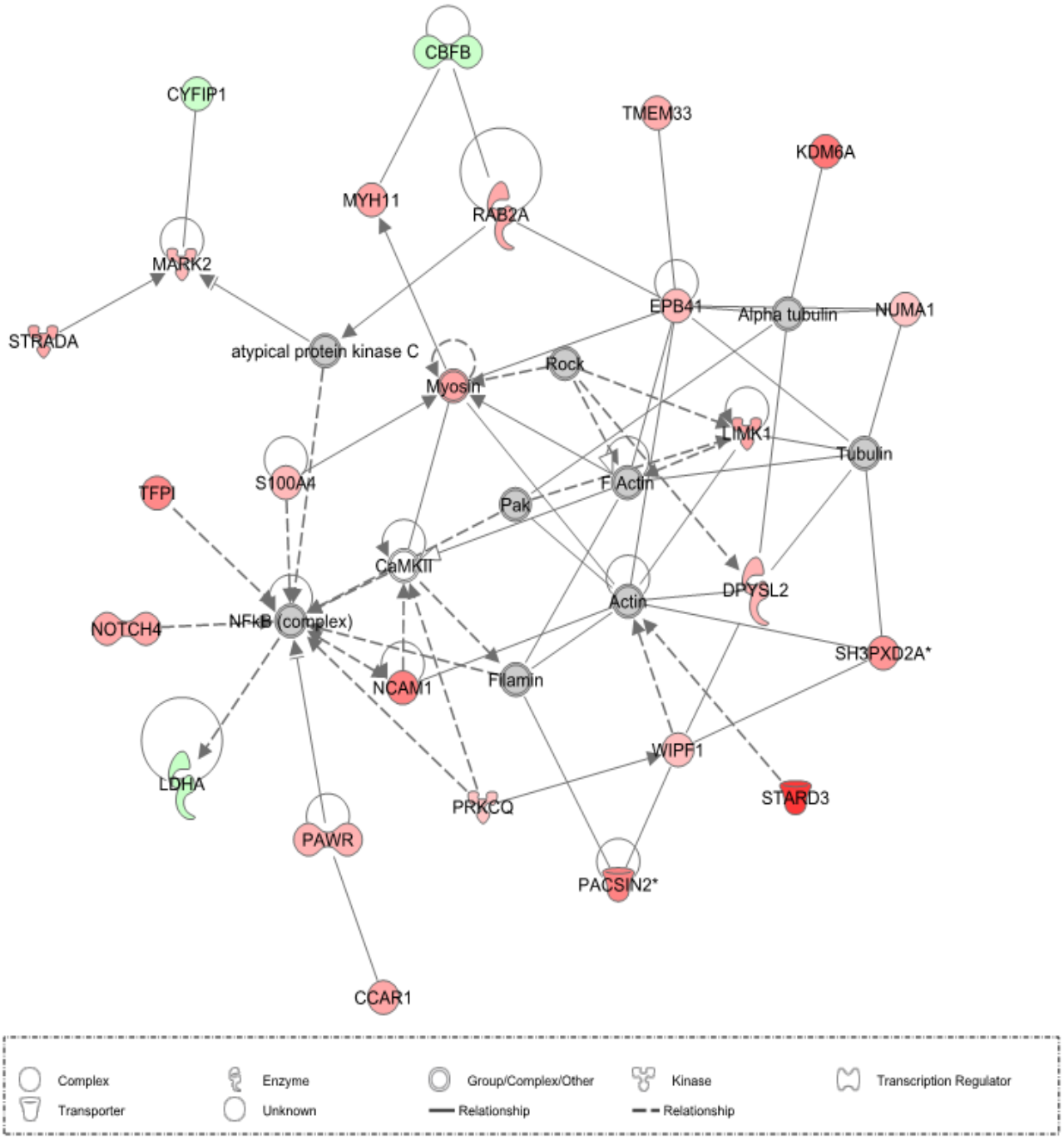

Figura 25: Genes modulados no adeocarcinoma pancreáticos estão envolcidos em movimento celular, expressão gênica e tráfico de células imunes. *Ingenuity Pathway Analysis* foi utilizado para identificar redes de interação gênica enriquecidas ao longo do conjunto de 221 codificadores de proteínas diferencialmente expressos em amostras de adenocarcinoma pancreático. A rede mais enriquecida (*p-value* < 10^{-41}) é composta por 24 genes diferencialmente expressos medidos pelo experimento de microarranjo. Vermelho indica alta expressão e verde, baixa expressão em tecidos tumorais em relação à pancreatite crônica e tecidos não tumorais pancreáticos.

4.6 LncRNAs intrônicos associados à metástase pancreática são expressos na orientação senso ou antissenso em relação ao mRNA (codificador de proteína).

Devido à falta de informação na literatura, fomos verificar mais a fundo a estrutura dos lncRNAs que mapeavam em *loci* gênicos associados a via de sinalização de Mapks e genes relacionados à apoptose. Para isso foram realizados experimentos de transcrição reversa fita específica para determinar se a orientação dos transcritos era senso ou antissenso em relação ao mRNA. Esta abordagem é importante para identificar se os transcritos intrônicos identificados possam ser provenientes de um processo pós-transcricional como *splicing* ou unidade independente de transcrição.

A transcrição reversa fita específica foi realizada para os 11 transcritos intrônicos que mapeavam em *loci* gênicos associados a via de sinalização de Mapks e genes relacionados à apoptose: *ARRB1*, *ATF2*, *MAPK1*, *TGFβR2*, *MAP2K5*, *MAP3K1*, *MAP3K14*, *PPP3CB*, *RAPGF2*, *DAPK1* e *PTEN*. Esses experimentos foram realizados utilizando RNA total isolados de amostras de tumores pancreáticos de pacientes (*pool*, n=3) e linhagem celular pancreática MIA PaCa-2. Dez dos 11 transcritos analisados mostraram evidências de serem transcritos na mesma orientação (senso) em relação ao mRNA codificador de proteína nas amostras de tumores pancreáticos e linhagem celular. Entretanto, para 3 desses 10 transcritos foram encontradas também transcrição na orientação antissenso em relação ao mRNA somente em amostras de tumores pancreáticos (*ATF2*, *TGFβR2* e *MAP3K1*). Interessantemente, um transcrito com orientação antissenso ao codificador de proteína no *locus* do *PPP3CB* foi identificado tanto em MIA PaCa-2 como em amostras tumorais primárias (Figura 26).

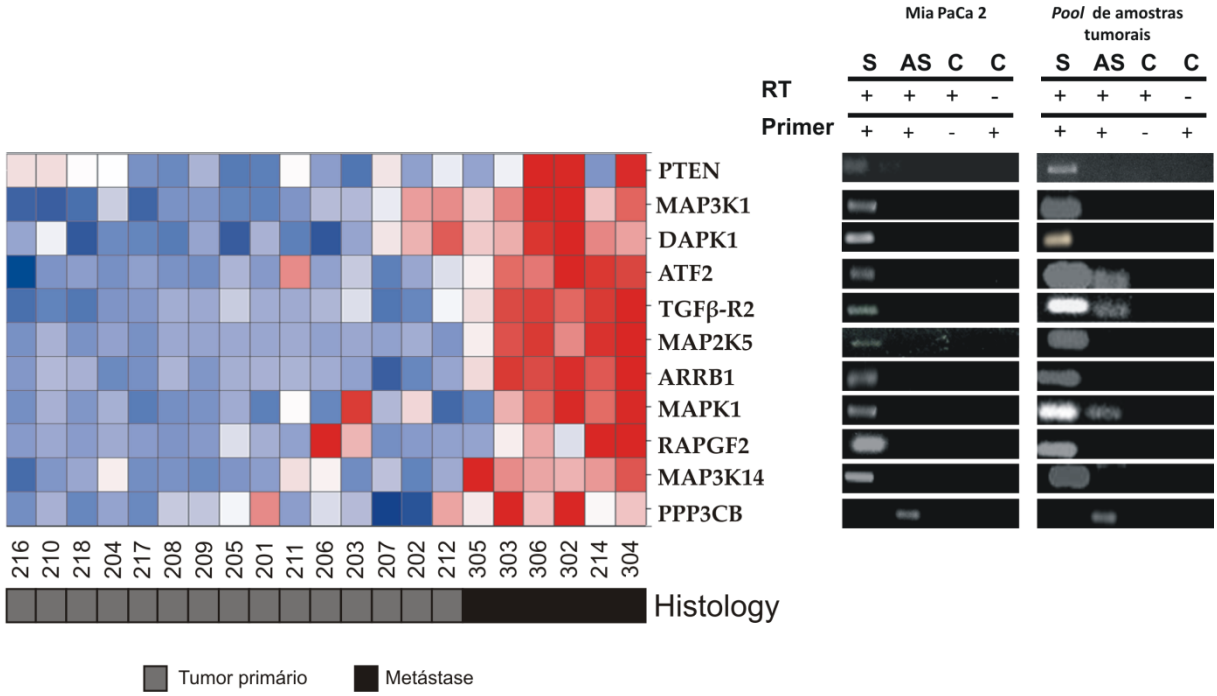

Figura 26: LncRNAs desregulados em metástase pancreática mapeiam em regiões intrônicas de genes associados com a via de sinalização das Mapks e/ou relacionados à apoptose. Orientação transcricional de onze transcritos intrônicos *ARRB1*, *ATF2*, *MAPK1*, *TGFβR2*, *MAP2K5*, *MAP3K1*, *MAP3K14*, *PPP3CB*, *RAPGF2*, *DAPK1* e *PTEN* foram investigados por transcrição reversa fita específica seguida de *PCR*. Para cada gene foi medida a transcrição senso (S) e antissenso (AS). Controles realizados sem *primers* foram realizados para verificar a presença de *self-anneling* durante a transcrição (RT+, Primer -). Controles para detectar contaminação de DNA genômica foram realizados com a ausência da transcriptase reversa (RT-, Primer +). O nível de expressão de cada gene está representado pelo número de desvios-padrões acima (vermelho) ou abaixo (azul) do valor médio de expressão do transcrito entre todas as amostras

A abundância relativa dos onze transcritos intrônicos foi avaliada por PCR quantitativo (qPCR) em amostras isoladas de pacientes que possuíam tumor primário e metástase. Algumas dessas amostras haviam sido utilizadas para a análise de microarranjos e 2 amostras de pacientes com metástase e 4 amostras de tumores primários utilizadas foram adquiridas de novos pacientes. Inicialmente, foram realizadas medidas para os 11 transcritos utilizando apenas um subconjunto de 3 amostras em cada uma das classes (tumor primário e

metástase) devido à limitação de quantidade de RNA total obtido de cada uma das amostras. Dos onze transcritos analisados, sete apresentaram uma mudança de expressão similar (aumento de expressão em amostras metastáticas) apontada pelos experimentos de microarranjos. Entretanto, essas mudanças de expressão não foram significativas devido ao pequeno conjunto de amostras e grande variabilidade de medida entre as amostras de pacientes.

Uma vez que a quantidade de RNA total das amostras era limitante, foram escolhidos 3 transcritos intrônicos para validação com um maior número de amostras, sendo um com orientação antissenso (*PPP3CB*) e dois com a mesma orientação (*MAP3K14* e *DAPK1*) relativa ao codificador de proteína. Como mostrado na Figura 27, os 3 transcritos intrônicos lncRNAs mostraram um aumento estatisticamente significante em amostras metastáticas em relação aos tumores primários.

Para investigar a coexpressão entre os transcritos intrônicos e os respectivos codificadores de proteína no mesmo *locus*, foram realizadas medidas de correlação de *Pearson* utilizando o nível de expressão medido em todas as amostras utilizadas. Somente o transcrito originado do *locus Map3k14* apresentou uma alta correlação de *Pearson* ($r = 0.78$, p-valor < 0.013), sugerindo que o transcrito intrônico senso possa ser originado do processamento do pré-mRNA do seu respectivo codificador de proteína (olhar a Discussão para maiores detalhes). Nenhuma correlação entre os codificadores de proteína e os transcritos intrônicos foi observada para os outros dois *loci* gênicos ($p > 0.05$), deixando a possibilidade desses transcritos intrônicos originados do *loci PPP3CB* (antissenso) e *DAPK1* (senso) serem unidades transcricionais independentes.

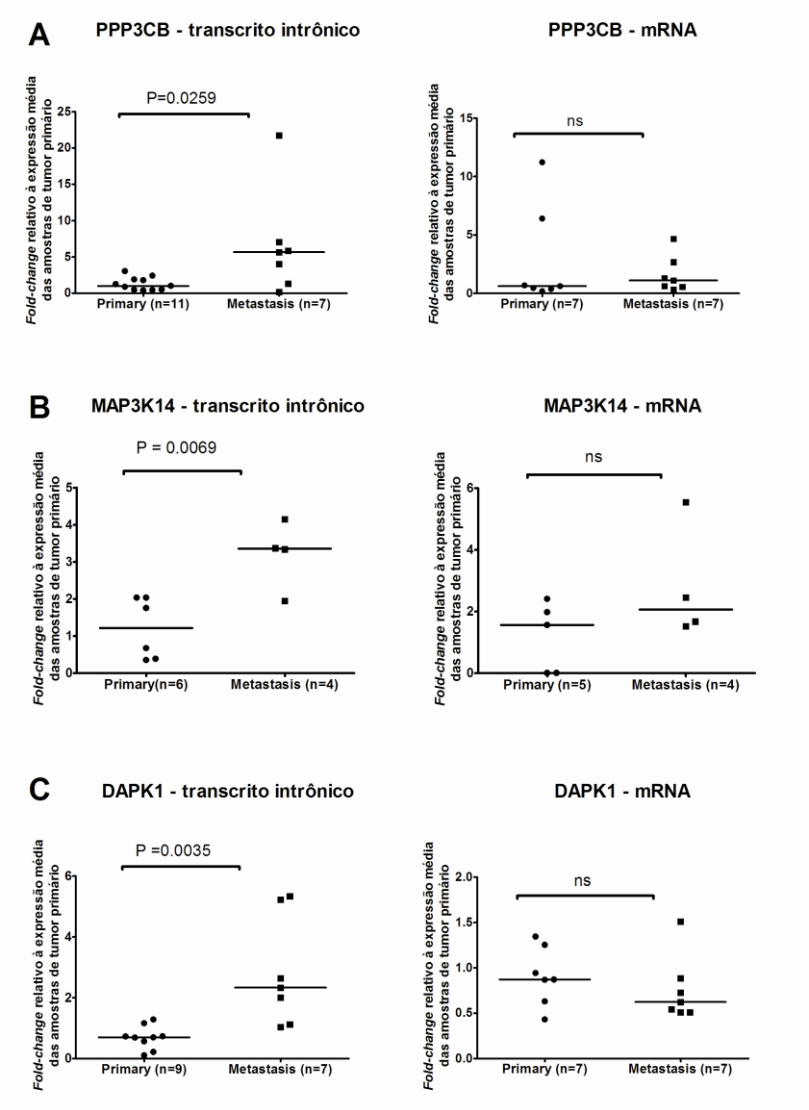

Figura 27: Mudanças no nível de expressão de lncRNAs intrônicos que mapeiam nos *loci* gênicos *PPP3CB*, *MAP3K14* e *DAPK1* e dos respectivos codificadores de proteína mRNAs. Níveis relativos de lncRNAs intrônicos e os respectivos codificadores de proteínas correspondentes aos *locus* gênicos de **PPP3CB (A)**, **MAP3K14 (B)** e **DAPK1 (C)** foram determinados em amostras clínicas de pacientes com tumores primários (círculos) e mestástase (quadrados). Para cadatranscrito, o número de amostras utilizadas está indicada no eixo X. Os resultados estão expressos em *Fold Change* relativo à expressão média dasamostras de tumores primários. As barras representam o valor da mediana de cada grupo. O valor de expressão de cada amostra foi calculado em triplicata e apenas o valor médio está representado no gráfico (círculo ou quadrado). O gene *HMBS* foi utilizado como controle endógeno para normalizar entre as amostras de pacientes. Todos os transcritos intrônicos (painéis esquerdo) apresentaram diferença de expressão estatisticamente significantes entre amostras de tumores primários e metástases ($P < 0.05$, teste-t não pareado) e o *p-value* está representado acima. Os transcritos codificadores de proteínas (painéis da direita) não apresentaram diferença de expressão significativa (ns).

4.7 LncRNAs associados à metástase pancreática apresentam tamanho médio de aproximadamente 1kb.

Dando continuidade aos resultados obtidos anteriormente, experimentos de RACE (*Rapid amplification of cDNA ends*) foram realizados para determinar a sequência e estimar o possível tamanho dos transcritos que foram validados por qPCR quantitativo. Somente os transcritos intrônicos senso provenientes dos *loci* de *MAP3K14* e *DAPK1* e o transcrito antissenso originado no *locus PPP3CB* foram analisados. O experimento foi realizado utilizando uma biblioteca comercial "*ready-to-use*" (MARATHON, Clontech) composta por cDNAs de tecido pancreático humano. Os resultados obtidos pela técnica de *RACE* apresentaram transcritos mais longos do que os mostrados pelas *ESTs* (cerca de 300bp) (Figura 28Figura 29Figura 30). Os tamanhos referentes aos transcritos intrônicos ncRNA senso mapeados nos *loci MAP3K14* e *DAPK1* é de ~1,5kb e ~1,1kb, respectivamente e o tamanho referente aos transcritos intrônico antissenso mapeado no *locus PPP3CB* é de ~1,4kb. As Figura 28Figura 29Figura 30 mostram um esquema representativo dos transcritos. Como é possível observar, todos os transcritos se localizam no primeiro intron das isoformas mais longas representadas em cada *locus*. Interessante notar que todos os transcritos foram identificados em outras células e o transcrito senso *MAP3K14* apresenta quantidade relevante de marcas de cromatina H3K4me1 e H3K27Ac. As modificações de cromatina H3K4me3 e H3K27Ac são associadas ao início de transcrição de RNAs conhecidos (Wang et al. 2008) e H3K4me1 estão associados a 20% dos enhancers (Wang et al. 2008). Segundo trabalho de Creyghton (Creyghton et al. 2010), a marca de H3K27ac juntamente com H3k4me1 identificam regiões de *enhancers* ativas.

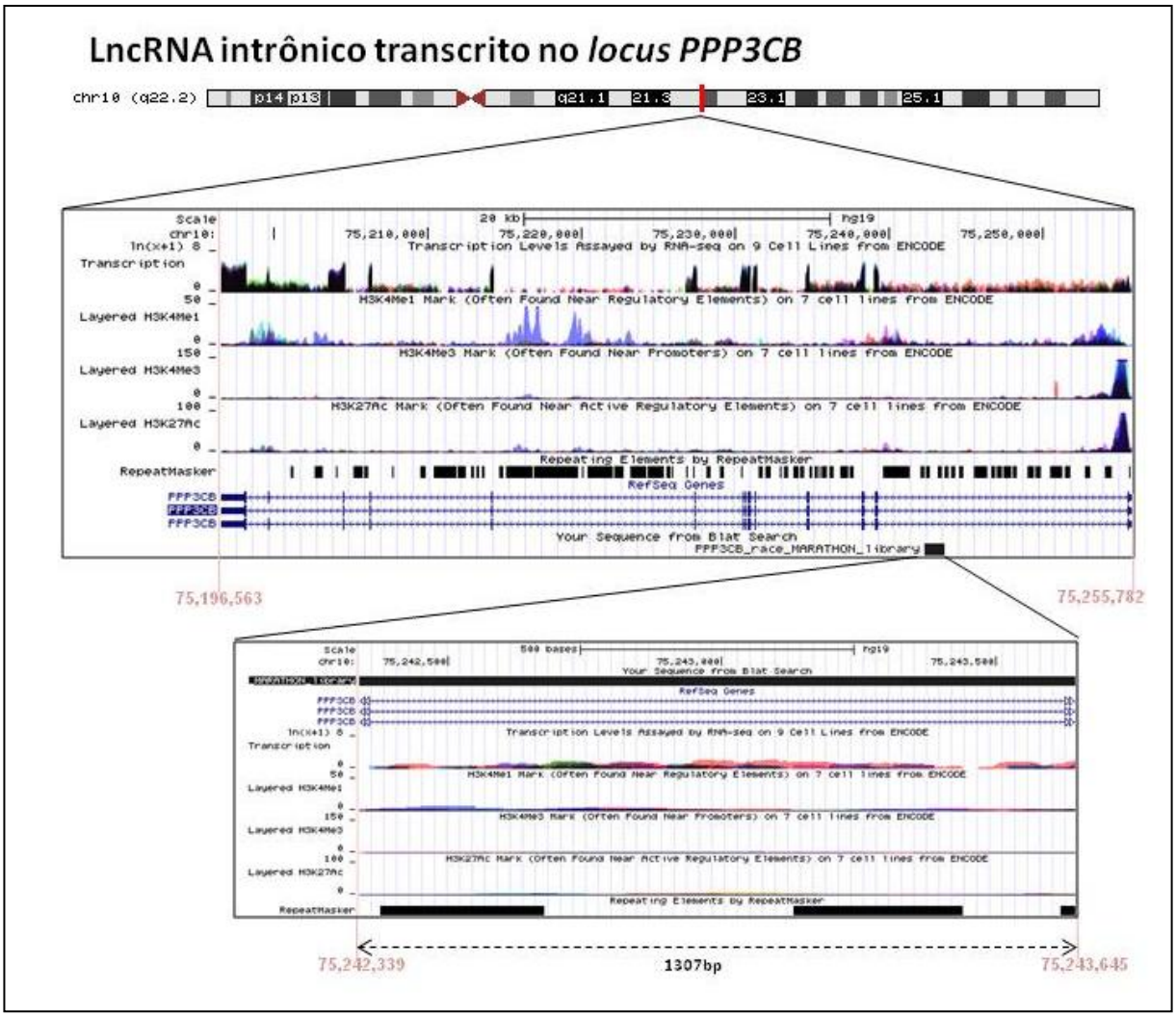

Figura 28: Mapeamento genômico do lncRNA antissenso *PPP3CB* e seu respectivo *locus*. O diagrama acima representa o *locus* gênico do gene. As coordenadas iniciais e finais estão representadas em rosa. As marcas de cromatina representadas são H3K4me1, muito associada à *enhancers*, H3K4me3, associada às regiões promotoras ativas e H3K27Ac, associadas à transcrição ativa e região promotora. Abordando 7 linhagens (GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK e NHFL) nenhuma proveniente de tecido pancreático. O *track* de transcrição é proveniente de RNA-seq *paried-end* (tamanho de 200nt e 75bp) *Illumina Genome Analyzer*, utilizando 9 linhagens (as sete descritas anteriormente e HeLaS3 e HepG2). Regiões repetitivas estão representadas em preto. O transcrito foi estendido apresentando tamanho 1307bp.

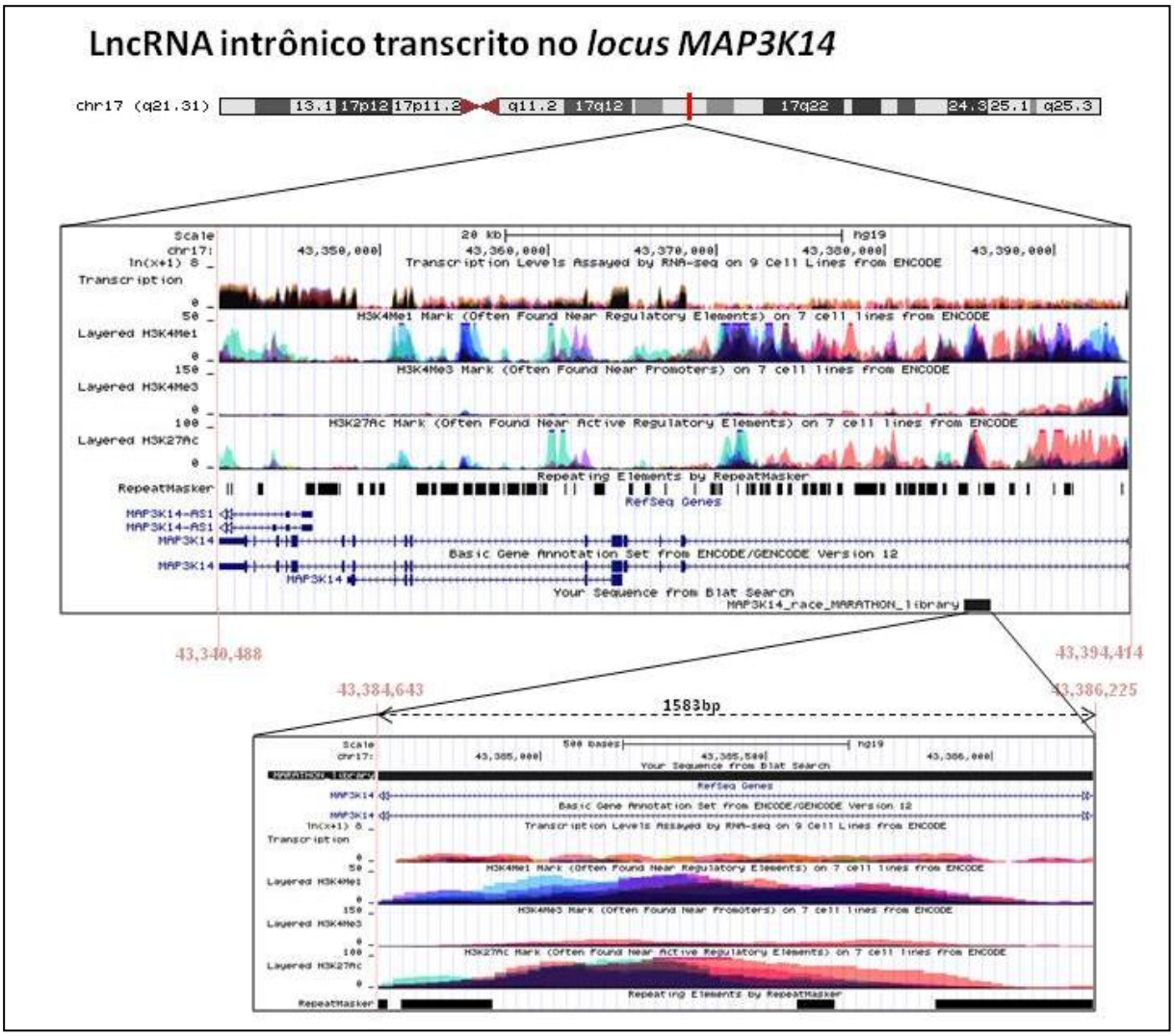

Figura 29: Mapeamento genômico do lncRNA intrônico senso *MAP3K14* e seu respectivo *locus*. O diagrama acima representa o *locus* gênico do gene. As coordenadas iniciais e finais estão representadas em rosa. As marcas de cromatina representadas são H3K4me1, muito associada à *enhancers*, H3K4me3, associada às regiões promotoras ativas e H3K27Ac, associadas à transcrição ativa e região promotora. Abordando 7 linhagens (GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK e NHFL) nenhuma proveniente de tecido pancreático. O *track* de transcrição é proveniente de RNA-seq *paried-end* (tamanho de 200nt e 75bp) *Illumina Genome Analyzer*, utilizando 9 linhagens (as sete descritas anteriormente e HeLaS3 e HepG2). Regiões repetitivas estão representadas em preto. O transcrito foi estendido apresentando tamanho 1583bp.

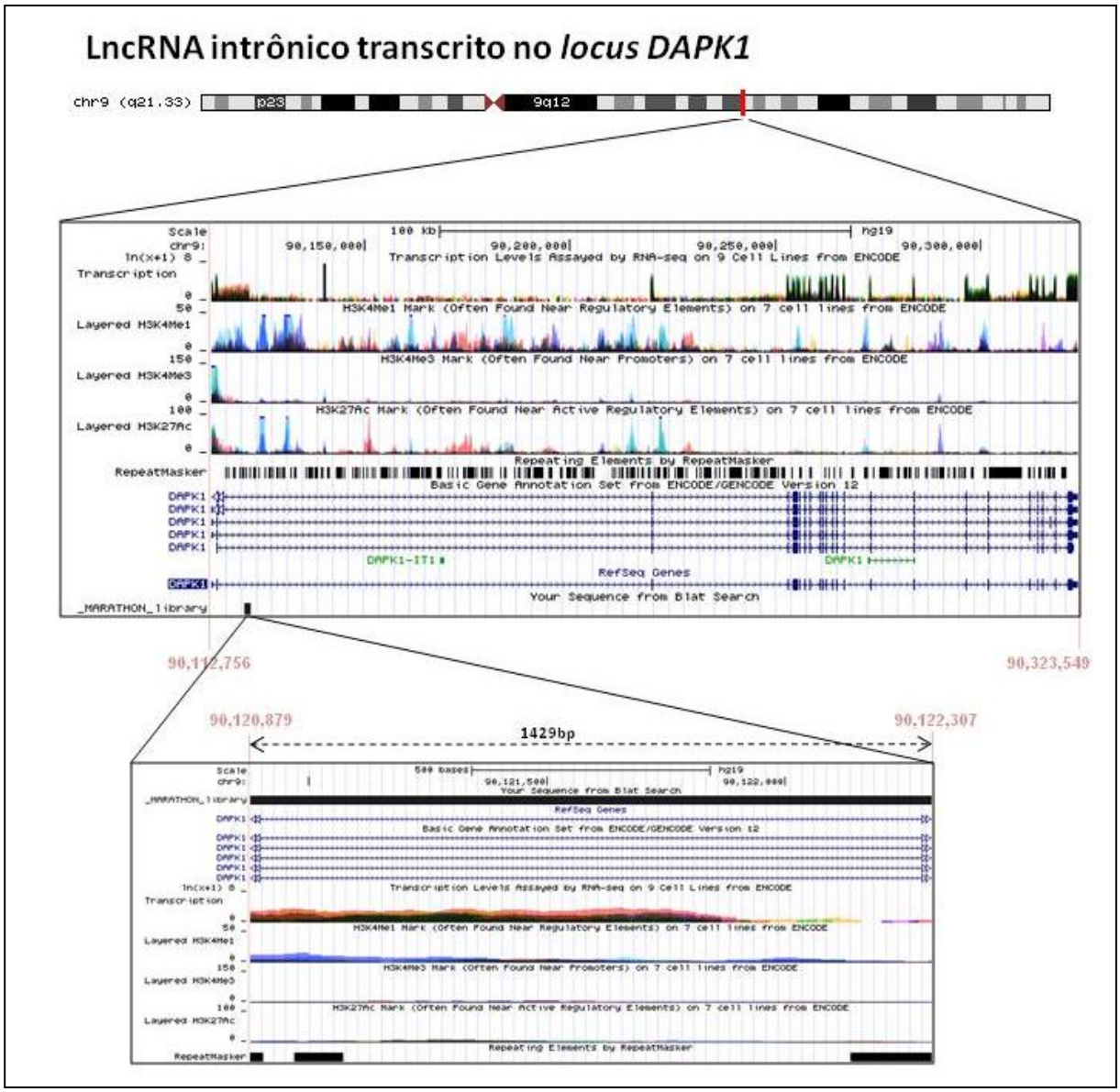

Figura 30: Mapeamento genômico do lncRNA intrônico senso *DAPK1* e seu respectivo *locus*. O diagrama acima representa o *locus* gênico do gene. As coordenadas iniciais e finais estão representadas em rosa. As marcas de cromatina representadas são H3K4me1, muito associada à *enhancers*, H3K4me3, associada às regiões promotoras ativas e H3K27Ac, associadas à transcrição ativa e região promotora. Abordando 7 linhagens (GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK e NHFL) nenhuma proveniente de tecido pancreático. O *track* de transcrição é proveniente de RNA-seq *paried-end* (tamanho de 200nt e 75bp) *Illumina Genome Analyzer*, utilizando 9 linhagens (as sete descritas anteriormente e HeLaS3 e HepG2). Regiões repetitivas estão representadas em preto. O transcrito foi estendido apresentando tamanho 1429bp.

A fim de tentar confimar o tamanho estendido por RACE, foi realizado um experimento de *northern blot* com sondas complementares ao transcrito antisenso intrônico expresso no *locus PPP3CB*. Este experimento teve como objetivo detectar o transcrito endógeno nas células Mia PaCa 2 para determinar o possível tamanho do transcrito. Como o transcrito intrônico é pouco expresso foi utilizado mRNA ao invés de RNA total na preparação do cDNA com o intuito de aumentar a representação relativa do transcrito. Assim, foram utilizados aproximadamente 20µg de mRNA a partir de 1,5mg de RNA total de células Mia PaCa 2. Ainda para tentar aumentar a sensibilidade da detecção, foram utilizadas na mesma hibridização 3 sondas que mapeavam em 3 regiões diferentes do transcrito para que aumentar o sinal. A Figura 31 mostra o resultado de *northern blot* para o transcrito *PPP3CB* após aproximadamente 2 semanas de exposição.

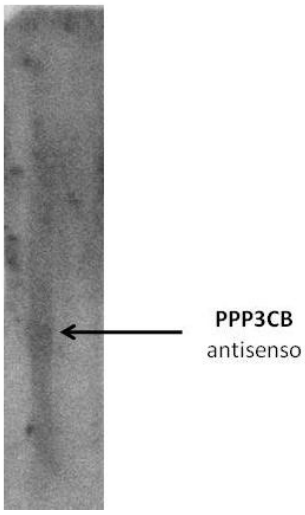

Figura 31: *Northern blot* para detecção do transcrito antisenso *PPP3CB*. Foi detectada uma banda com tamanho compatível com um transcrito de 1500-1800nt. A partir de 3 sondas marcadas radioativamente de 20µg RNA enriquecidos em poli(A)+ com duas semanas de exposição.

4.8 LncRNAs senso *DAPK1* possui tamanho de 1050pb confirmados por RT-PCR fita-específica em MIA PaCa-2 e é expresso em Capan-1.

Os ensaios de RACE mostraram que este transcrito possuía tamanho aproximado de 1,1kb. Para confirmar que pelo menos este tamanho era plausível foram desenhados *primers* nas pontas do transcrito fora de regiões repetitivas, permitindo a amplificação de um transcrito de 1050pb. Realizando RT-PCR fita-específica com RNA total da linhagem MIA PaCa-2, foi detectada uma banda no tamanho esperado. Esta banda foi clonada e sequenciada, mostrando que o transcrito encontrado de fato era o transcrito senso intrônico *DAPK1* (Figura 32).

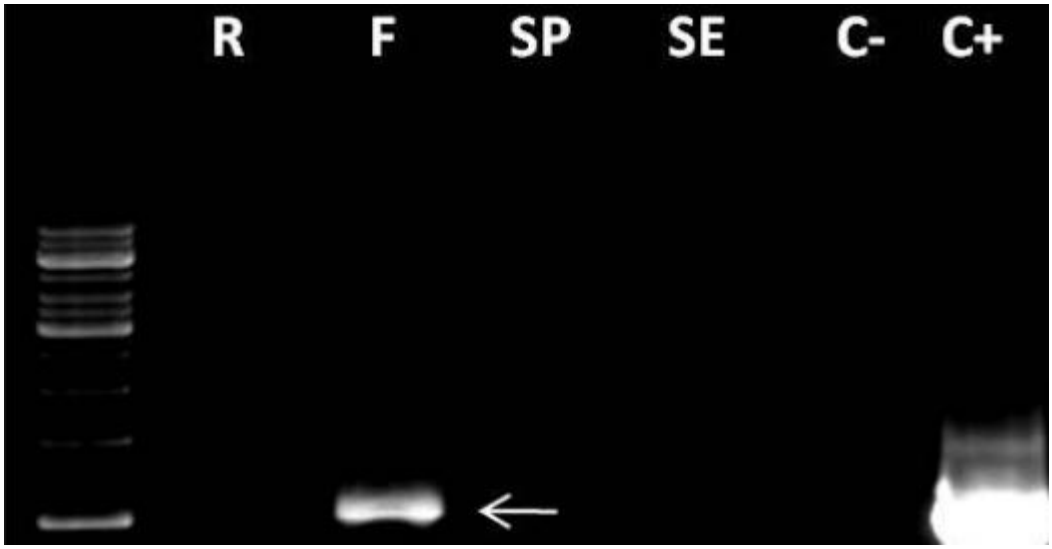

Figura 32: RT-PCR fita-específica do transcrito intrônico *DAPK-1*. Foi observado que o transcrito DAPK-1 possui pelo menos 1050pb, como visto por RACE. A seta indica esse tamanho esperado. Os *primers* detectam a orientação antissenso (AS), senso (S), controle sem *primers* para verificar a evidência de *self-annealing* do RNA (SP) e controle sem enzima na RT para detecção de contaminação genômica (SE). C- = controle da PCR sem enzima; C+= controle positivo da PCR com gDNA extraído de células HeLa.

Os resultados anteriores apontavam para o fato de que o transcrito intrônico *DAPK-1* apresentava orientação senso quando analisado em RNA total de MIA PaCa-2 e no *pool* de

tumores utilizados na análise (Figura 26). Com o intuito de verificar a orientação deste transcrito em outras linhagens pancreáticas foi utilizada a técnica de RT-PCR fita-específica com amostras de RNA total das seguintes linhagens PANC-1, AsPC-1 e Capan-1 (Figura 33).

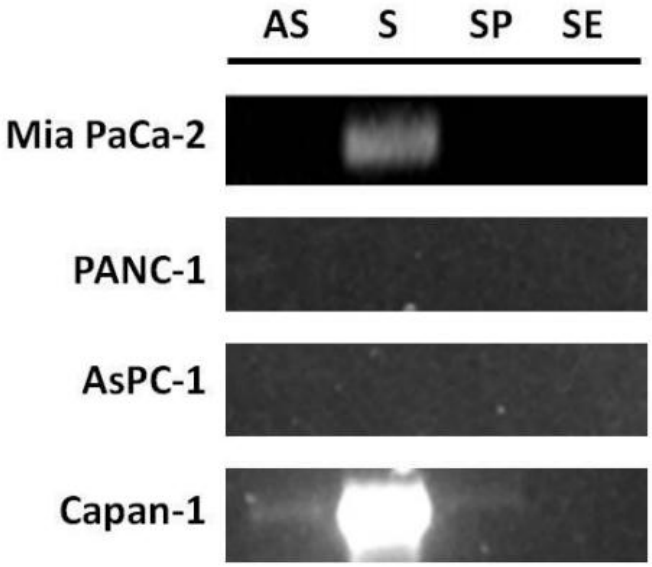

Figura 33: Orientação do transcrito intrônico *DAPK-1* em linhagens pancreáticas. Transcrição reversa realizada com *primers* específicos apresentados que detectam a orientação antissenso (AS), senso (S), controle sem *primers* para verificar a evidência de *self-annealing* do RNA (SP) e controle sem enzima na RT para detecção de contaminação genômica (SE).

Somente as linhagens MIA PaCa-2 e Capan-1 apresentaram a expressão do lncRNA intrônico *DAPK-1*, mas a expressão desse transcrito não é observada nas células PANC-1 e AsPC-1. Como é possível notar na Figura 33, o transcrito intrônico é expresso na orientação senso relativa ao seu correspondente codificador de proteína. Interessantemente, é possível notar que a maior expressão deste transcrito ocorre na linhagem celular metastática Capan-1 em relação à linhagem de tumor primário MIA PaCa-2, este resultado ainda foi confirmado por qPCR como apresentado na Figura 34.

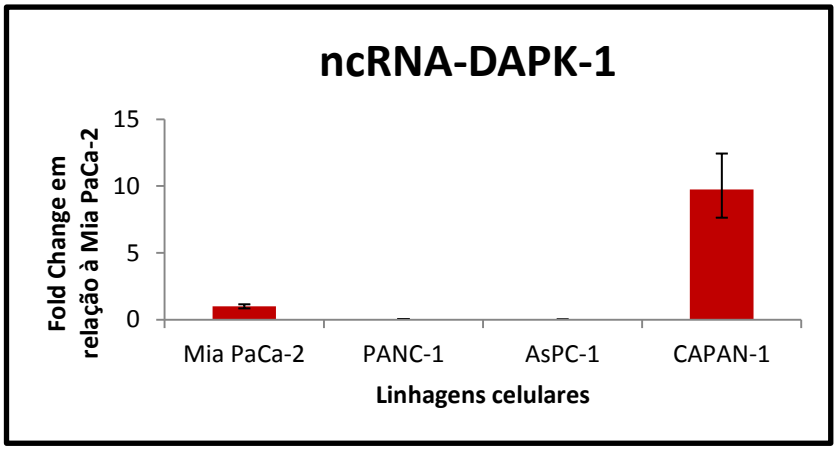

Figura 34: qRT-PCR para observar a expressão de RNA nas linhagens celulares pancreáticas. Expressão do transcrito intrônico não codificador *DAPK-1*. Ambos os transcritos foram normalizados a partir da expressão do gene *HMBS* e o *Fold Change* calculado em relação ao nível de expressão apresentado na linhagem MIA PaCa-2. As barras de erros apresentadas são referentes às réplicas técnicas

4.9 LncRNA senso *DAPK1* possui região promotora e pode estar associado a presença das proteínas de isoforma curta e longa.

Pelo fato de se tratar de um transcrito com orientação senso é muito provável que este seja originado da transcrição reversa do pre-mRNA ou introns *lariats* do que uma unidade transcricional independente. Após a confirmação do tamanho, procuramos por regiões a montante do possível início de transcrição utilizando o programa TSSG disponível no site softberry com o link descrito a seguir (http://linux1.softberry.com/berry.phtml?topic=tssg&group=programs&subgroup=promoter). O programa mostrou uma região promotora contendo TATAbox. Esta mesma região foi analisada utilizando o site Alibaba2.1 (http://www.gene-regulation.com/pub/programs/alibaba2 /index.html) para procurar por possíveis fatores de transcrição que poderiam se ligar na sequência predita como promotora. O site prediz sítios de ligação de fatores de transcrição a partir da construção de matrizes utilizando o TRANSFAC. Utilizando esta ferramenta de bioinformática foram localizados dois sítios para a ligação de

TBP (TATA-binding protein), que são proteínas que se ligam em regiões tipo TATAbox. Um dos sítios sobrepõe exatamente a região predita pelo primeiro programa TSSG (Figura 35).

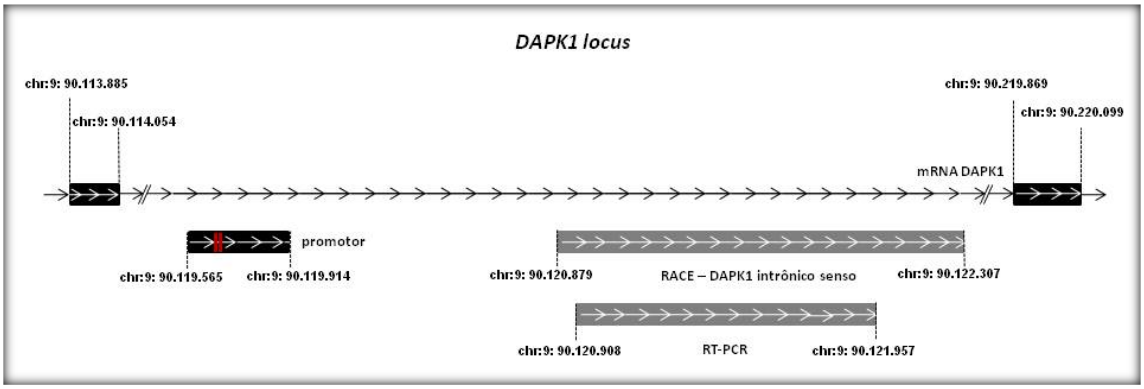

Figura 35: Diagrama representativo do transcrito intrônico DAPK1 e seu respectivo locus. O diagrama mostra o transcrito intrônico DAPK1 (cinza) estendido por RACE (~1,4kb) e confirmação por RT-PCR fita específica (1050bp). O mRNA está representado acima com regiões intrônicas (setas pretas) e exônicas (caixas pretas e setas brancas). A possível região promotora predita está representada à montante do transcrito intrônico por caixa preta com setas brancas, as duas barras vermelhas representam os dois sítios de ligação preditos para ligação de TBP (*TATA-binding protein*). As coordenadas genômicas estão mostradas em relação à montagem hg19.

Além destas análises, também foi feita uma busca para identificar marcas de cromatina H3K4me3 associadas a regiões promotoras (Guttman et al. ; Mikkelsen et al. 2007). Foram analisadas 14 linhagens celulares, nenhuma relacionada ao tecido pancreático, e identificamos que algumas linhagens mostraram uma marcação de trimetilação da cromatina muito discreta em relação ao threshold apresentado pelo genome browser, somente a linhagem GM 12878 (células humanas linfobastóides) apresentou marcação de H3K4me3 com alta intensidade. Desta forma, há fortes indícios que existe uma promotora que regula a transcrição do transcrito senso DAPK1. Após encontrar evidências bioinformáticas de que de fato o transcrito *DAPK1* poderia possuir uma região promotora, foi realizado um ensaio de promotor com o gene repórter da *luciferase* (Dual-Luciferase Reporter Assay System, Promega) em

Mia PaCa2 como descrito na seção 3.13.1 em materiais e métodos. As células foram transfectadas com várias construções diferentes de vetores sendo pGL3-basic (controle negativo), pRL-SV40 (controle positivo), pGL3 com a sequência referente ao promotor e pGL3 com a sequência invertida. Como mostrado na figura 7, a possível sequência promotora possui aproximadamente 50% da atividade promotora do vetor pRL-SV40, considerado um promotor forte. Este resultado juntamente com os resultados anteriores sugere que o transcrito intrônico *DAPK1* é um transcrito independente do codificador de proteína e que é regulado por um promotor, sugerindo que este transcrito seja de fato funcional.

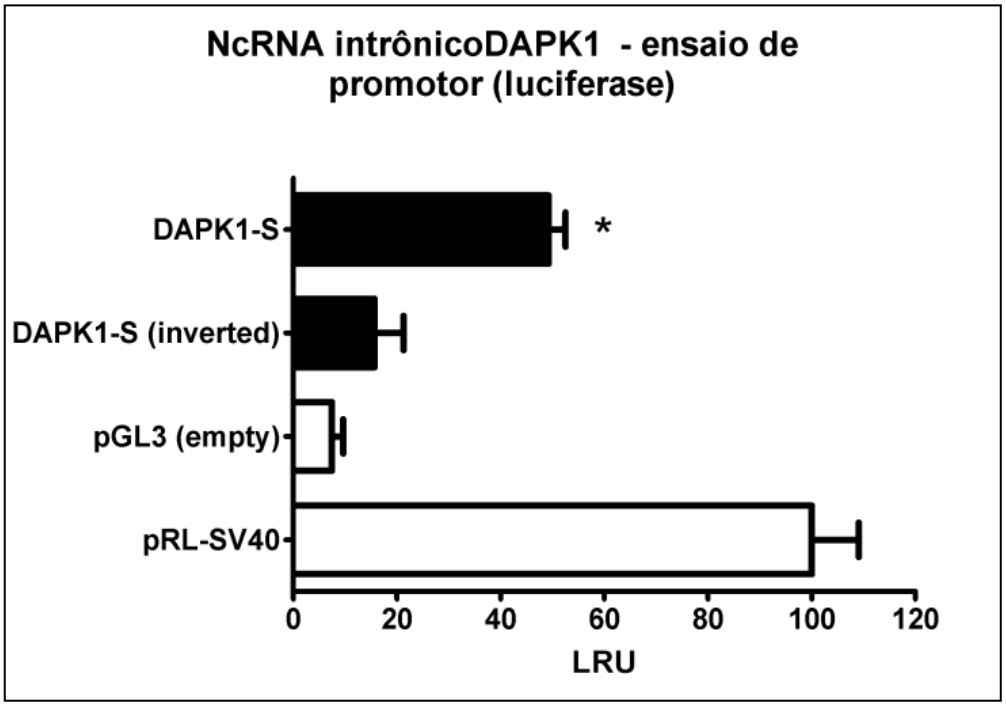

Figura 36: Ensaio de promotor (luciferase). O gráfico apresenta as unidades relativas de luminescência (LRU) da luciferase de firefly em relação à luciferase renilla. A primeira representa a atividade promotora da sequência em teste e a segunda o controle de transfecção na célula. As sequências testadas estão representadas em barras pretas e os controles negativo (pGL3-empty) e positivo (pRL-SV40) estão representadas por barras brancas. As barras de erros são referentes a três réplicas biológicas contendo 3 réplicas técnicas cada uma. O asterisco representa medidas que possuem atividade promotora diferente do controle negativo com significância estatística (ANOVA, seguido de teste de *Tukey*, *p-value* < 0.05).

Embora ainda seja necessário confirmar o tamanho total do ncRNA DAPK-1 os resultados obtidos indicam que este transcrito seja gerado a partir de um promotor próprio, com orientação senso e de forma independente do mRNA produzido no *locus* DAPK-1.

4.10 A expressão do lncRNAs senso *DAPK1* não parece estar correlacionada com as isoformas de mRNAs dos correspondentes codificadores de proteína do *locus DAPK-1*

O transcrito lncRNA *DAPK-1* apesar de ser senso foi escolhido para ser estudado, pois havia evidências de que a sua transcrição fosse independente pelo fato de não apresentar correlação de expressão com o seu respectivo codificador de proteína em amostras clínicas de pacientes (r=0.2, correlação de Pearson). Além disso, a identificação de atividade promotora à montante do transcrito intrônico, suporta ainda mais a ideia de que este lncRNA seja uma unidade transcricional independente. O transcrito não-codificador escolhido para ser estudado mapeia em uma região intrônica do gene DAPK-1, que codifica a proteína com função pró-apotótica Death-associated protein kinase 1 (Kogel et al. 2001). Essa proteína possui 1430 aminoácidos e possui importantes domínios para sua função, entre eles o "*Death Domain*" responsável por sua característica pró-apoptótica. Esse domínio se localiza na região C terminal, como mostrado no trabalho de Stevens e colaboradores (Stevens et al. 2009).

A busca deste *locus* gênico no site do UCSC (http://genome.ucsc.edu/) revelou que haviam outras isoformas codificadoras descritas para este *locus*. Interessantemente, a isoforma mais curta, que não possui o *Death Domain* foi identificada no banco de dados de ESTs a partir de tecidos tumorais como mostrado na Figura 37.

Figura 37: Isoformas correspondentes ao gene *DAPK-1*. Figra obtida no site do *genome browser*. As isoformas mais curtas depositadas no banco de mRNAs, e que não possuem o *Death Domain*, estão reresentadas por setas ao lado indicando a origem das amostras onde foram identificadas. Nota-se que isoformas menores foram identificadas em amostras tumorais de dois diferentes tecidos (tumor de língua e melanoma).

Uma vez que estava descrito na literatura a possibilidade de RNAs intrônicos antissenso alterarem o balanço de isoformas do codificador de proteína sem alterar substancialmente a expressão do respectivo *locus* como no caso do transcrito *Saf* (Yan et al. 2005), foi investigada a possibilidade do transcrito não codificador proveniente do *locus DAPK-1* estivesse associado com o balanço das isoformas curta e longa, não alterando a expressão do total do gene *DAPK-1*, mas sim de suas isoformas.

Para testar essa hipótese, foram realizadas qPCRs em amostras clínicas de pacientes com tumores primários e metástases pancreáticas utilizando dois pares de *primers*. O primeiro par de *primers* se anelava em todas as isoformas codificadoras apresentadas no *loci* gênico *DAPK-1* e o outro anelava apenas na isoforma longa que apresentava o *Death-Domain*. Com este ensaio, esperava-se observar uma maior quantidade de expressão de isoformas longas (pró-apoptótica) nas amostras de tumores primários em relação às amostras metastáticas.

Além de verificar a presença da isoforma longa transcrito codificador *dapk-1*, também foi a analisada a possibilidade de correlação entre a expressão do lncRNA e da isoforma longa.

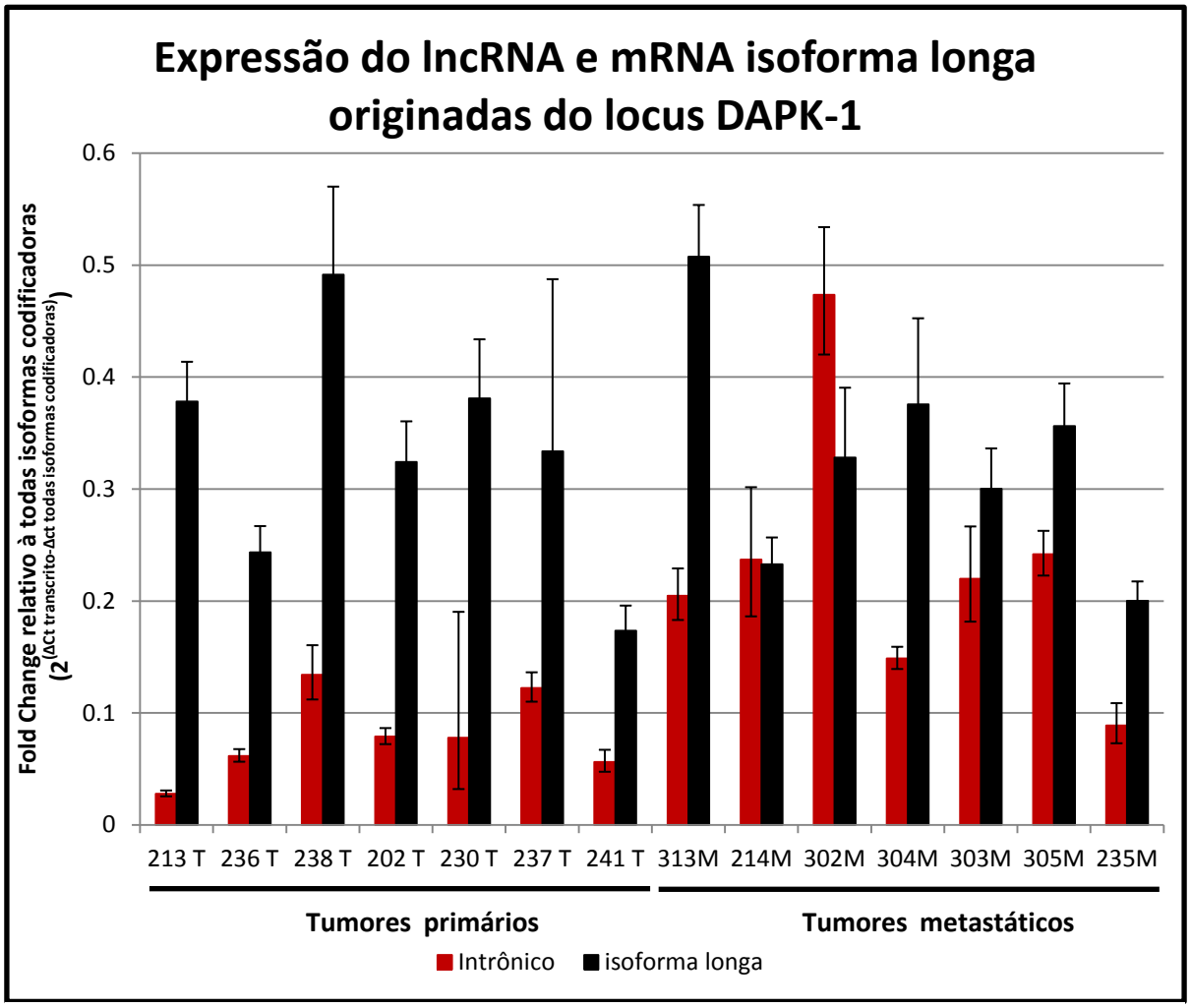

Figura 38: Relação das isoformas longas e do transcrito intrônico *DAPK-1* em relação a todas as isoformas. Níveis relativos de lncRNAs intrônicos e da isoforma longa associada ao *locus DAPk-1* foram determinados em amostras clínicas de pacientes com tumores primários e metástase pancreáticas. Os resultados estão expressos em *Fold Change* relativo à expressão de todas as isoformas codificadoras detectadas em cada amostra. O valor de expressão de cada amostra foi calculado em triplicata e apenas o valor médio está representado no gráfico. O gene *HMBS* foi utilizado como controle endógeno para normalizar entre as amostras de pacientes. A isoforma codificadora longa não apresentou diferença de expressão estatisticamente significantes entre amostras de tumores primários e metástases e também não houve correlação de expressão entre a isoforma longa e lncRNA (r=0.12, correlação de Pearson).

Não foram observados maiores valores de expressão do transcrito que codifica a isoforma longa de DAPK1 nos tumores primários em relação às metástases como era esperado. Realizando-se o teste-t obtêm-se um p-valor > 0.05, o que indica que não há uma validação estatística nessa análise. Estes resultados sugerem que o transcrito intrônico da DAPK1 não possui correlação com a expressão do RNA codificador de proteína.

4.11 A expressão do lncRNA senso DAPK1 parece estar correlacionada com a dapk1 isoforma longa em linhagens pancreáticas MIA PaCa 2 e Capan-1.

Uma vez que o transcrito intrônico senso *DAPK-1* parecia não agir em *cis*, ou seja não afetar a expressão do seu respectivo codificador de proténa, uma estratégia para revelar a sua função seria fazer o silenciamento do transcrito para identificar os efeitos fenotípicos causados. Para isso, a primeira análise a ser feita foi medir a quantidade relativa do nível de expressão dos RNAs intrônicos e da isoforma longa em linhagens pancreáticas.

O resultado abaixo mostra uma correlação positiva entre o transcrito intrônico não codificador e o mRNA correspondente à isoforma longa (r = 0.999, p-valor = 0.001). Este padrão se mostrou contraditório ao que havia sido visto nos pacientes, entretanto isso pode ser explicado, parcialmente, pelo fato de amostras de pacientes possuírem uma população heterogênea de células, enquanto que as linhagens tratam de uma população específica de célula pancreática. Esta falta de correlação do nível de expressão de pacientes e linhagens celulares também foi mostrada em artigo publicado recentemente associando *linc (long intergenic ncRNAs) HOTAIR* com adenocarcinoma pancreático (Kim et al. 2012). Apesar das evidências de que o transcrito intrônico ncRNA *DAPK-1* era proveniente de uma unidade transcricional independente, foi realizado um qPCR com *primers* desenhados em outras regiões intrônicas que não possuíam evidências transcricionais com o intuito de medir a

expressão de pré-mRNAs (Figura 40). Como pode ser observado o transcrito intrônico possui uma maior expressão que o seu respectivo mRNA em linhagens Capan-1 e MIA PaCa 2.

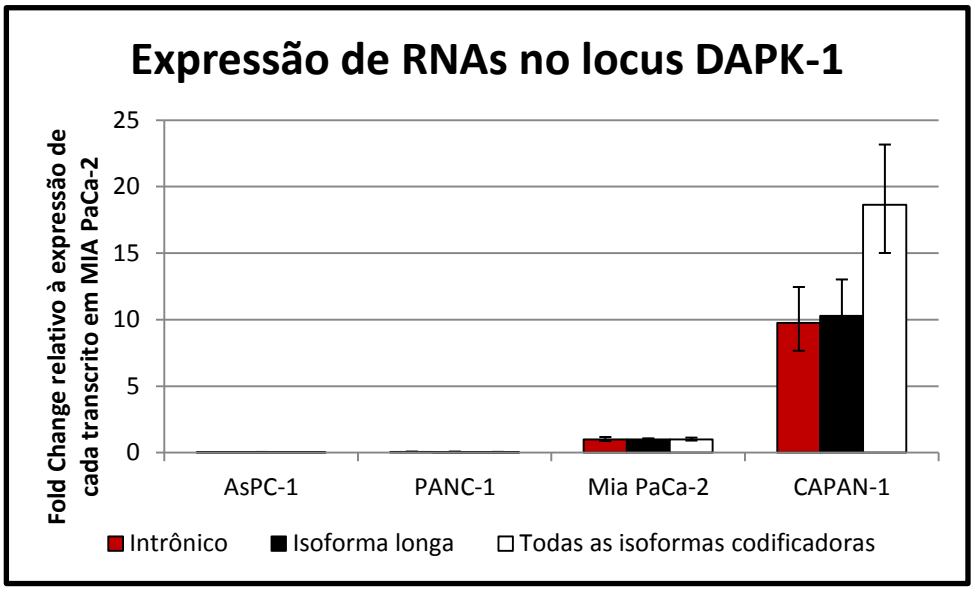

Figura 39: Expressão relativa de RNAs no *locus DAPK-1* em linhagens celulares pancreáticas. Expressão relativa dos RNAs intrônicos (barras vermelhas), RNA mensageiro de DAPK-1, com *primers* que detectavam o RNA originado da isoforma longa dessa proteína (barras pretas) e todas as isoformas codificadoras (barras brancas). Os transcritos foram normalizados a partir da expressão do gene *HMBS* e o *Fold Change* calculado em relação ao nível de expressão apresentado na linhagem MIA PaCa-2. As barras de erros apresentadas são referentes às réplicas técnicas.

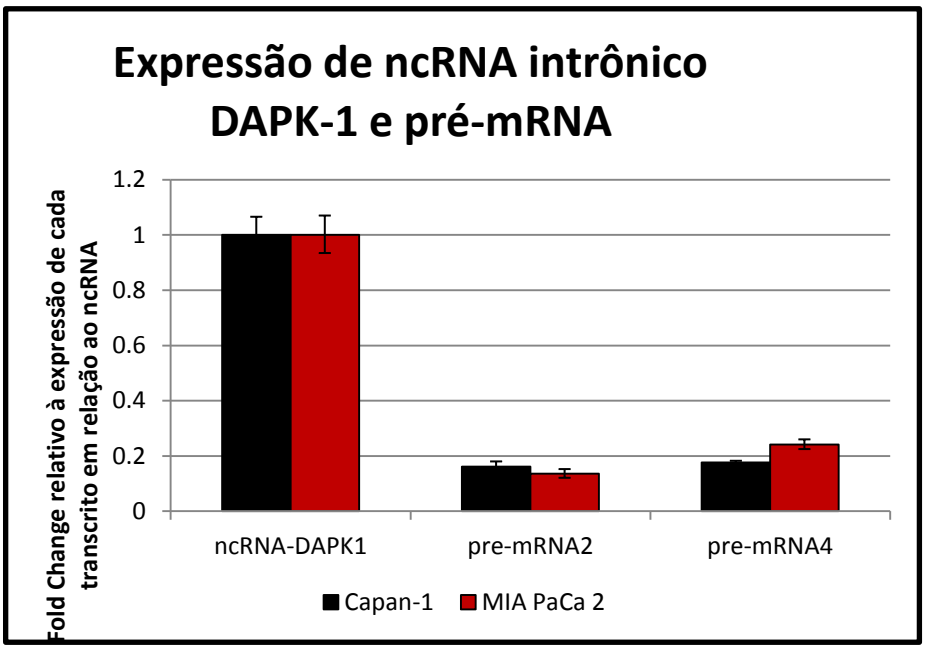

Figura 40: Expressão relativa de pré-mRNAs no locus DAPK-1 em linhagens celulares pancreáticas Capan-1 e MIA PaCa 2. Expressão dos pré-RNAs (intron 2 e intron 4) em relação ao ncRNA intrônico *DAPK-1* em linhagens Capan-1 (barras pretas) e MIA PaCa 2 (barras vermelhas). Os transcritos foram normalizados a partir da expressão do gene *HMBS* e o *Fold Change* calculado em relação ao nível de expressão apresentado do ncRNA intrônico *DAPK-1*. As barras de erros apresentadas são referentes às réplicas técnicas.

A partir desses resultados promissores começou-se a trabalhar na pardronização para o silenciamento do transcrito intrônico *DAPK1* com oligos antissensos modificados (fosfotioato e O-Metil), a fim de saber se este transcrito está correlacionado com a isoforma codificadora de proteína longa. O tratamento com oligo antissenso foi realizado em 24 e 48 hs. Oligos antissenso para o silenciamento dos transcritos de actina e MALAT-1 foram utilizados como controle positivo do tratamento. Após o tratamento, foi realizado um *Real-time* RT-PCR com *primers* para verificar o silenciamento desses transcritos, bem como o silenciamento do transcrito codificador da proteína DAPK-1 (DAPK1_exonic). Os resultados desse experimento encontram-se na Figura 41.

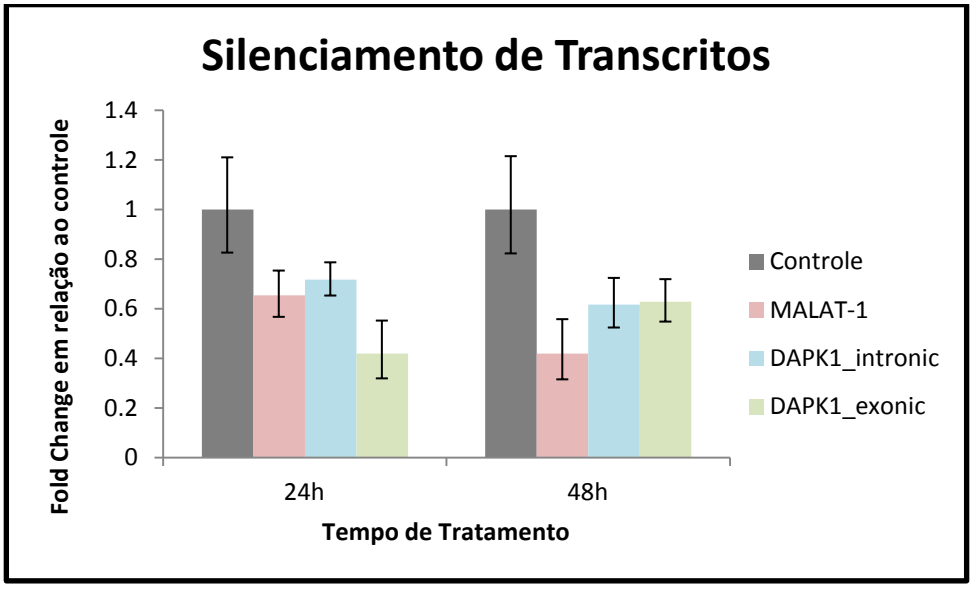

Figura 41: Silenciamento de transcritos utilizando oligos antisenso modificados. Nesse experimento as células Capan-1 foram tratadas 24h e 48h com oligos para silenciar o *MALAT-1*, o ncRNA *DAPK-1* e também células tratadas com Lipofectamina, mas sem a adição de oligos antisenso modificados. O *Real Time* RT-PCR foi utilizado com *primers* que detectavam o transcrito *MALAT-1*, o ncRNA *DAPK-1*, o codificador de proteínas DAPK-1_exonic. O *fold change* foi calculado em relação ao controle. As barras de erros são referentes às réplicas técnicas.

Nesse experimento, foi possível observar a diminuição do transcrito intrônico não codificador DAPK-1 assim como a diminuição da expressão do transcrito codificador da proteína DAPK-1 e do codificador da mini proteína s-DAPK-1.

Com esse experimento não é possível concluir se o silenciamento do transcrito codificador é devido ao anelamento do oligo antisenso no pré-mRNA promovendo a diminuição da expressão ou se trata de um efeito secundário após o silenciamento do ncRNA *DAPK-1*. A fim de verificar se de fato está havendo um silenciamento do pre-mRNA, foi realizado outro *Real Time* RT-PCR com RNA das amostras de 48h e com *primers* que detectavam diferentes íntrons sem evidência de transcrição, para comparar a expressão destes com o ncRNA *DAPK-1* (Figura 42).

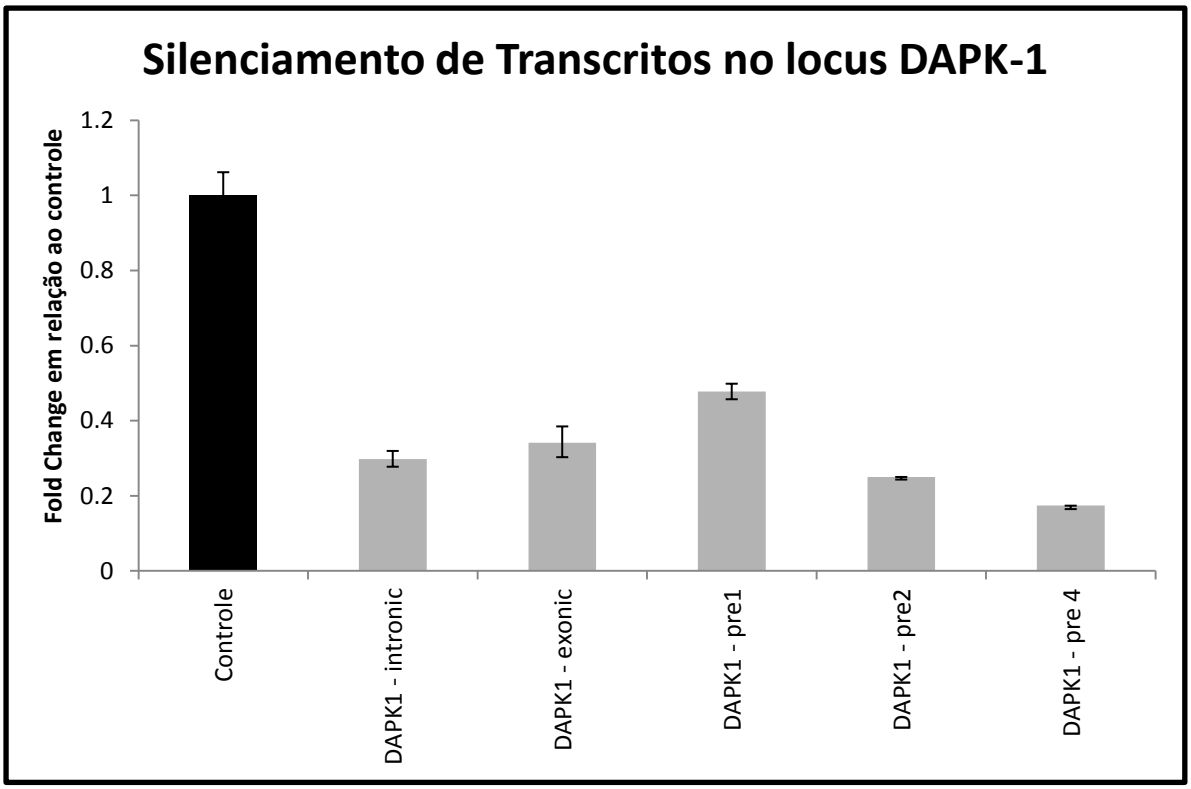

Figura 42: Silenciamento de transcritos utilizando oligos antissenso modificados. Utilizando RNA extraídos das células tratadas em 48h foi feito novo *Real Time* RT-PCR com *primers* que detectavam o ncRNA *DAPK-1*, s-DAPK-1, DAPK-1_exonic, além de três regiões do pre-mRNA que não possuem evidências de transcrição: DAPK1_pre1, DAPK1_pre2 e DAPK1_pre3. O *Fold Change* foi calculado em relação ao nível de expressão desses transcritos na amostra controle. As barras de erros são correspondentes às réplicas técnicas.

Como pode ser observado no gráfico acima, a taxa de silenciamento dos transcritos analisados é praticamente a mesma em todos, sugerindo que de fato está ocorrendo anelamento dos oligos no pre-mRNA, uma vez que foi detectada a diminuição da expressão tanto do ncRNA *DAPK-1*, do codificador de proteínas, quanto das três regiões de pré-mRNA analisadas.

5 Discussão

Neste trabalho, nós investigamos o perfil de expressão de amostras clínicas de pacientes que possuíam adenocarcinoma pancreático usando uma plataforma de microarranjos customizada. Esta plataforma, que já havia sido utilizada em outros estudos do nosso grupo (Reis et al. 2004; Louro et al. 2007; Louro et al. 2008), era composta por sondas que interrogavam possíveis RNAs não codificadores longos mapeados em regiões intrônicas e intergênicas do genoma humano e uma coleção de mRNAs que já haviam sido associados ao câncer na literatura. Comparando o perfil de expressão de 38 amostras clínicas de tecido pancreático com 4 diferentes histologias (15 adenocarcinoma primário, 9 tecidos não tumoral adjacente ao tumor, 8 pancreatites crônicas e 6 tecido metastáticos), nós detectamos nos 4 tipos de tecidos transcritos não codificadores que mapeavam em regiões intrônicas e intergênicas em menor proporção do que os codificadores de proteína (Figura 10). Esta constatação deve-se ao fato de que transcritos não codificadores são expressos em níveis menores que os codificadores de proteína (Nakaya et al. 2007; Guttman et al. 2009; Cabili et al. 2011; Rinn and Chang 2012), logo, grande parte desses transcritos devem ter sido excluídos durante as etapas de filtragem dos dados gerados pelos experimentos de microarranjos.

Foi visto que a maior parte dos transcritos intrônicos e intergênicos expressos nos tecidos pancreáticos apresentavam pequeno ou nenhum potencial codificador (96% do total), utilizando o CPC (Figura 14). Mesmo os ncRNAs que apresentaram ORFs preditas pelo programa *ORFfinder* (embutido no programa CPC) possuíam baixo potencial codificador devido ao pequeno tamanho das ORFs (Figura 14A), sugerindo que a maior parte desses RNAs não codifique para proteína. Esta hipótese é reforçada com o trabalho de Banfai e colaboradores (Banfai et al. 2012), no qual estima-se que, aproximadamente, 92% de lncRNAs definidos pela anotação do GENCODE v7 não são traduzidos. Para verificar se os

ncRNAs expressos nos tecidos pancreáticos poderiam possuir tamanho maior do que o apresentado pelas ESTs, foi realizada uma comparação com sequências de *contigs* formadas pela montagem de ESTs com *splicing* criada pelo nosso próprio grupo (Nakaya et al. 2007). Esta análise mostrou que esses transcritos apresentavam um tamanho médio de 779 nt (Figura 15), sendo maiores do que as ESTs depositadas na plataforma de microarranjo, mostrando que essas sequências representam apenas parte de transcritos mais longos. Para nos assegurar que os transcritos intergênicos não fossem UTRs (*Untranslated Regions*) não anotadas, foi calculada a distância dos transcritos intergênicos em relação às pontas UTRs. Foi visto que a maior parte dos transcritos (81% do total) estava situada a mais de 1kb de distância das pontas UTRs, sugerindo que a maior parte destes transcritos são de fato intergênicos do que UTRs não caracterizadas.

Com a conclusão do projeto ENCODE, iniciado em 2007 (Birney et al. 2007) e concluído em 2012 (Dunham et al. 2012), fica claro que os lncRNAs podem exercer diversas funções celulares por meio de diversos mecanismos (Mattick 2009; Wilusz et al. 2009; Hung and Chang 2012; Reis and Verjovski-Almeida 2012; Rinn and Chang 2012). Entretanto, muitos grupos ainda discutem a possibilidade de que a fração de transcritos não codificadores identificados seja resultado da transcrição espúria devido à atividade da RNA polimerase II nas regiões de cromatina abertas ou, ainda, seriam segmentos intrônicos processados de mRNAs (Struhl 2007; van Bakel et al. 2010). No nosso trabalho, não é possível distinguir se os transcritos intrônicos longos medidos são provenientes de: **i**- regiões intrônicas (*lariats*) resultantes do processamento de pré-mRNA ou **ii**- unidades transcricionais independentes localizadas dentro de uma região intrônica do genoma. Para minimizar a captura de introns *lariats*, nós focamos na fração RNAs poliadenilados utilizando iniciadores de oligo-dT para a transcrição reversa, diminuindo as chances de capturar RNAs não poliadenilados provenientes do processo de *splicing* de mRNAs. Além disso, a identificação de um subconjunto de

transcritos que mapeiam em regiões intrônicas e possuem expressão maior que o seu respectivo codificador de proteína, sugere que esta fração de transcritos intrônicos se acumule e não sejam rapidamente degradados como os introns *lariats* (Cheng and Menees 2011). Ainda, foi realizada uma série de análises para obter evidências adicionais que suportam a hipótese de que os transcritos não codificadores longos e intrônicos e intergênicos são, de fato, transcritos *bona fide*. Além disso, a identificação de 96 transcritos (29%) na base de dados do GENCODE v12, evidencia que esses transcritos são não codificadores longos e ainda suporta a hipótese de que os outros lncRNAs identificados como expressos em tecido pancreático também possam estar representados na base de dados do GENECODE futuramente.

Com o intuito de reforçar as evidências de que estes transcritos não codificadores possam ser transcritos *bona fide*, foi utilizada a abordagem para a busca de elementos conservados na sequência de DNA (*phastConElements*), pequenas sequências que tem sido conservada entre 44 espécies de vertebrados. Nesta análise ainda utilizamos a base de dados disponível para *phastConElements* de apenas mamíferos ou primatas. Apesar de a fração conservada ser pequena, foi observado um enriquecimento estatisticamente significante (teste de Fisher, p-valor < 0.05) de elementos conservados quando comparados ao grupo randômico de sequências. Este resultado sugere que pelo menos uma fração desses lncRNAs sob seleção positiva nas espécies de vertebrados e, portanto, podem ser biologicamente funcional. Entretanto, para os transcritos que não possuíam elementos conservados no DNA determinados pelo *phastConElements*, a ausência de conservação de sequência não deve ser levada como evidência de irrelevância funcional, uma vez que é conhecido que os lncRNAs bem caracterizados (*MALAT-1* e *XIST*) são pouco conservados globalmente (Pang et al. 2006). O lncRNA *HOTAIR*, que já é conhecido por agir em *trans*, também apresenta baixa conservação de sequência com considerável conservação de estrutura em 11 mamíferos (He et

al. 2011). Esta afirmação ainda pode ser reforçada por um elegante estudo feito por Zhang e colaboradores, o qual demostrou que a substituição de uma região particular do transcrito MEG3 com uma sequencia não homóloga, que preservava a estrutura secundária do transcrito, não teve nenhum efeito na ativação de p53, sugerindo que a estrutura secundária é importante (Zhang et al. 2010). Ainda, o estudo realizado por Ulitsky e colaboradores (Ulitsky et al. 2011), dois lincRNAs (*cyrano* e *megamind*), que possuíam apenas pequenos trechos de alta conservação, foram silenciados em embriões de *zebrafish* produzindo fenótipos defeituosos em relação ao sistema nervoso. Entretanto, a expressão dos ortólogos de camundongo ou humano, em embriões de *zebrafish* com lincRNA silenciado, foi capaz de resgatar o fenótipo normal para parte dos embriões [35-60%], indicando que apesar de não haver uma grande conservação de sequência, há conservação na função desses RNAs não codificadores. Acredita-se que os lncRNAs possuem taxa de mutação maior do que a seleção neutra, atrelando papéis funcionais que contribuem para diferenças biológicas de espécies (Kutter et al. 2012).

Como proposto por Walshietl e colaboradores (Washietl et al. 2005), predições de estrutura secundária podem levar a descoberta de novos RNAs não codificadores. Apenas uma pequena fração de lncRNAs expressos em pâncreas (15%, i.e. 49/335) mostrou predição de domínios de estrutura secundária estável que poderiam ser importantes para o próprio processamento ou função biológica. Entretanto, sabe-se que o algoritmo RNAz usado possui a limitação de utilizar a conservação da sequência primária como um dos parâmetros para a busca de estrutura secundária (Gruber et al. 2007). Como já discutido no parágrafo acima, os lncRNAs são globalmente pouco conservados em relação a sua sequência primária, logo se utilizarmos outros tipos de algoritmos que deem menos importância à conservação da sequência, provavelmente, poderíamos ter encontrado um número maior de estruturas secundárias estáveis.

Na literatura já está bem documentado que pequenos RNAs regulatórios podem ser gerados do processamento de RNAs não codificadores longos transcritos de regiões intrônicas e intergênicas do genoma (Kapranov et al. 2007; Fejes-Toth et al. 2009; Wilusz et al. 2009). Para identificar qual seria a fração de lncRNAs expressos em tecidos pancreáticos poderiam originar pequenos RNAs regulatórios, as coordenadas genômicas dos lncRNAs foram comparadas com bancos de dados conhecidos de miRNAs e snoRNA (Lestrade and Weber 2006; Griffiths-Jones et al. 2008). Somente uma sobreposição discreta foi encontrada, indicando que os transcritos longos (intrônicos e intergênicos) não são, predominantemente, precursores conhecidos de microRNAs ou snoRNAs. Entretanto, ainda há a possibilidade que estes transcritos representem precursores de pequenos RNAs não caracterizados.

A fim de buscar evidências na literatura que suportem a existência dos transcritos longos intrônicos e intergênicos identificados neste trabalho, foi feita a busca de evidência transcricional em dados de RNA-seq disponíveis gerados a partir de nove tecidos distintos (Wang et al. 2008). Foi identificado que aproximadamente 80% dos transcritos não codificadores (intrônicos e intergênicos) detectados em tecidos pancreáticos foram detectados em pelo menos uma das nove bibliotecas analisadas (Figura 13). A maior parte dos transcritos identificados nas bibliotecas de RNA-seq foi detectada em apenas um tipo de tecido sem ser pancreático, ou nos nove tecidos abordados pela técnica de RNA-seq, além do pâncreas. Este comportamento entre os transcritos não codificadores identificados em tecidos pancreáticos indica a prevalência de subconjuntos de lncRNAs com perfil de expressão geral e tecido específico, respectivamente.

O perfil de expressão tecido específico para uma fração de RNA não codificadores, sugere que a sua expressão seja regulada (Nakaya et al. 2007; Mercer et al. 2008). Como proposto nos trabalhos de Mikkelsen (Mikkelsen et al. 2007) e Guttman (Guttman et al. 2009)

marcas de cromatina como a trimetilação da lisina 4 na histona 3 (H3K4me3) são frequentemente encontradas em regiões de RNAs transcritos pela RNA polimerase II. Na análise de transcritos não codificadores mapeados em regiões intrônicas, foi encontrado enriquecimento desta marca de cromatina (H3K4me3) na vizinhança de 2kb de distância do início de transcrição quando comparado com 100 grupos randômicos de sequências de DNA (KS, p-valor < 0.05). Um enriquecimento comparável foi encontrado para os mRNAs codificadores de proteína, sugerindo que a transcrição de mRNAs e lncRNAs intrônicos apresentam assinatura de cromatina similar na região promotora. Além disso, também foi observado um enriquecimento significante (KS, p-valor < 0.05) de CAGE *tags* próximo ao início de transcrição conhecido para os transcritos intrônicos expressos nos tecidos pancreáticos, corroborando com a hipótese que parte desses transcritos possa representar unidades transcricionais independentes. Uma vez que as bibliotecas de CAGE *tags* utilizadas não possuíam linhagens provenientes de tecidos pancreáticos, a quantidade de transcritos intrônicos ou intergênicos que possuem início de transcrição com CAP5' pode ainda estar subestimada.

Diferentemente dos mRNAs codificadores de proteínas, não foi encontrado enriquecimento significante para a presença de ilhas CpGs na vizinhança dos lncRNAs expressos em tecidos pancreáticos estudados neste trabalho. Baseado nesta observação, nós propomos que a metilação de ilhas CpGs canônicas não parece estar envolvida na regulação transcricional da maior parte dos lncRNAs intrônicos e intergênicos expressos em tecidos pancreáticos. Esta hipótese ainda é reforçada por Sati e colaboradores (Sati et al. 2012) que estudou quatro linhagens diferentes (h1, PBMC, Brain Fr e Brain Gr) e também não identificou densidade de metilação à montante do início de transcrição de lncRNAs, diferindo do padrão apresentado pelos mRNAs codificadores de proteína.

Diante de várias evidências de estrutura secundária, conservação e de regulação fica claro que pelo menos uma fração desses transcritos (intrônicos ou intergênicos) seja proveniente de unidade transcricional independente do que simplesmente ruído transcricional originado do inespecificidade da RNA polimerase II (Struhl 2007). Isto nos levou a investigar em maiores detalhes os níveis de expressão desses lncRNAs entre as amostras tumorais e não tumorais.

Diferenças de expressão de lncRNAs intrônicos em câncer de próstata e rim já foram documentadas pelo nosso próprio grupo (Reis et al. 2004; Brito et al. 2008). Neste trabalho, nós estendemos essas observações para adenocarcinoma pancreático, a fim de verificar se havia conjuntos de lncRNAs desregulados em amostras clínicas de pacientes com câncer de pâncreas. Comparando o perfil de expressão de tumores primários com amostra histologicamente não tumorais e pancreatite crônica, foi identificada uma assinatura de 147 transcritos correlacionados com tumor primário de pâncreas. Esta estratégica foi utilizada para favorecer o aparecimento de marcadores especificamente associados ao adenocarcinoma pancreático do que transcritos associados com componentes de células estromais, as quais estão presentes em células tumorais e pancreatite crônica (Gress et al. 1997; Logsdon et al. 2003). Apesar de morfologicamente parecidas, as células estromais presente em pacientes com pancreatite crônica e pacientes com tumor pancreático apresentam diferenças moleculares como visto por Pilarsky e colaboradores (Pilarsky et al. 2008). Entretanto, comparando os transcritos encontrados como diferencialmente expressos entre as células estromais de pancreatite crônica e tumor primário com a assinatura encontrada neste trabalho, não houve sobreposição dos genes encontrados, sugerindo que a maior parte do que foi visto na plataforma de microarranjo seja inerente ao tumor pancreático. Contudo, atualmente, a presença de células estromais no tumor pancreático vem atraindo a atenção de grupos de pesquisa, e está sendo apontada como uma das maiores responsáveis pela agressividade do

adenocarcinoma pancreático (Erkan et al. 2012; Kong et al. 2012). Como o objetivo desta análise era priorizar os genes associados ao tumor primário de pâncreas em relação ao tecido não tumoral não foi comparada a diferença de expressão de pancreatite crônica e amostras tumorais.

Devido à quantidade limitante de RNA remanescente dos pacientes, uma das opções para checar se a assinatura encontrada estava coerente com os dados da literatura foi investigar se existiam transcritos que já haviam sido identificados em outros trabalhos da literatura relacionados à expressão gênica de tumores pancreáticos. Apenas 23% dos mRNAs codificadores de proteína presentes na assinatura foram também identificados em outros estudos. Esta pequena sobreposição pode ser consequência de diferentes plataformas utilizadas e da heterogeneidade das amostras tumorais (Jarvinen et al. 2004). Entretanto, foi observada uma grande concordância (17/24, 71%) entre as diferenças de expressão encontradas na assinatura obtida pelo nosso trabalho e os perfis de expressão gênicos apresentados por outros grupos. Este fato contribui para dar suporte aos resultados obtidos e indica que a nosso grupo de amostras e metodologia foram aplicados corretamente.

Alguns dos genes já reportados na literatura como diferencialmente expressos no adenocarcinoma pancreático já haviam sido estudados como biomarcadores para este tipo de câncer (por exemplo S100A6 (Ohuchida et al. 2007), S100P (Deng et al. 2008), TIMP1 (Pan et al. 2011) e NF-κB (Weichert et al. 2007)). Em concordância com trabalhos anteriores (Lopez-Casas and Lopez-Fernandez 2010), a análise de enriquecimento de categorias na assinatura de expressão gênica associada ao adenocarcinoma pancreático apresentou o enriquecimento de genes envolvidos em adesão focal (Figura 20). Além disso, dentre os 16 genes mostrados na Tabela 6 associados às vias enriquecidas encontradas pela análise de GO, sete já tinham sido reportados como diferencialmente expressos em outros tipos de

malignidades associados ao tecido pancreático, como lesões pré-malignas PanIN e IPMN. O gene *VCL*, além de ter sido detecatado como aumentado em lesões pré-malignas PanIN-1B/2 (Prasad et al. 2005), também foi identificado com expressão aumentada em amostras de tumores pancreáticos, sendo considerado como um potencial oncogene (Loukopoulos et al. 2007). Interessantemente, este gene teve sua proteína detectada em urina de indivíduos saudáveis (Adachi et al. 2006), apontando para possíveis candidatos a serem estudados como marcadores moleculares futuramente.

O enriquecimento da via de adesão focal sugere que a alteração de genes codificadores de proteínas relacionados à conexão e sinalização da matriz extracelular tenha um papel importante na transformação maligna ou manutenção do adenocarcinoma pancreático. Utilizando o IPA, a rede mais enriquecida encontrada estava relacionada a genes associados às vias de movimento celular e interação célula-célula, reforçando a importância da via de adesão focal. O conjunto de genes encontrados como alterados no adenocarcinoma pancreático incluía a integrina beta 5 (*ITGB5*), a qual foi encontrada como aumentada nas amostras tumorais. A proteína Itgb5 já havia sido investigada como biomarcador em câncer de pulmão de células não pequenas (Dingemans et al. 2008) e é alvo da droga inibidora EMD121974, a qual está sob testes clínicos (Manegold et al. 2012). Desta forma, itgb5 é um potencial candidato para ser testado como novo alvo de drogas em câncer de pâncreas.

É importante ressaltar que uma fração (29%) dos 147 transcritos, que compunham a assinatura associada ao adenocarcinoma primário pancreático, era composta por lncRNAs mapeando em regiões intrônicas e intergênicas, sugerindo que os RNAs não codificadores poderiam exercer papéis relacionados à tumorigênese do câncer de pâncreas. Este resultado nos levou a investigar a existência de conjuntos de lncRNAs cm níveis de expressão alterados em amostras metastáticas. Atualmente, há um aumento de interesse na relação entre a

presença de lncRNAs e câncer. Alguns dos lncRNAs já foram identificados como alterados em cânceres como o *MALAT-1* por contribuir com a migração celular em câncer de bexiga (Ying et al. 2012) e pulmão (Tano et al. 2010), *HOTAIR*[20] já foi visto como um possível preditor para cancer de mama (Rinn et al. 2007) e também a expressão deste transcrito já foi descrita como aumentada em amostras de tumor pancreático em relação a amostras não tumorais (Kim et al. 2012). Outros trabalhos e revisões apontam vários lncRNAs correlacionados com outros tipos de cânceres (Prensner and Chinnaiyan 2011; Nie et al. 2012; Reis and Verjovski-Almeida 2012). Apesar do crescente interesse nos lncRNAs, a literatura apresentada de lncRNAs mapeados em regiões intrônicas ainda é muito escassa, o que suporta o nosso interesse em identificar esses transcritos no adenocarcinoma pancreático.

A assinatura associada à metástase pancreática identificou 355 transcritos diferenciamente expressos, os quais incluíam 220 codificadores de proteínas, 134 transcritos intrônicos e intergênicos e 6 lncRNAs conhecidos, alguns já haviam sido relacionados com o adenocarcinoma pancreático como mostrado na seção 4.5. Esta assinatura identificou codificadores de proteína correlacionados com metástases em outros tipos de câncer (Apêndice 6), apontando para potenciais candidatos interessantes para serem testados como novos alvos para tratamento de metástase no adenocarcinoma pancreático.

A fração significante de lncRNAs encontrada na assinatura de metástase (38% do total) sugere que a desregulação desses lncRNAs poderiam estar associados também com o processo metastático. A alteração de expressão de mRNAs de genes enriquecidos na via de *MAPK* quinase já foi descrita no carcinoma pancreático (Zhao et al. 2006; Campagna et al. 2008; Yachida and Iacobuzio-Donahue 2009). Neste trabalho, foram identificados nove

[20] *HOTAIR* presente em mamíferos com sequência pouco conservada (Rinn et al. 2007; He et al. 2011). O transcrito fica localizado entre os genes HOXC10 e HOXC11 e age em *trans* silenciando genes da família *HOXD* através do recrutamento de PRC2 (Rinn et al. 2007).

transcritos que mapeavam em regiões intrônicas de genes pertencentes à via de *MAPK* quinase na assinatura de metástase. Também foram observados alteração de expressão em *loci* gênicos associados à apoptose, incluindo 42 codificadores de proteína e 6 lncRNAs intrônicos, esta via foi uma das 12 descritas pelo trabalho de Jones e colaboradores (Jones et al. 2008) como geneticamente alterada no câncer de pâncreas. Estes resultados nos levaram a documentar melhor a natureza desses 11 transcritos mapeados em regiões intrônicas de genes de *loci* gênicos associados à via de *MAPK* e apoptose.

Para caracterizar melhor esses transcritos, o primeiro ensaio a ser realizado foi identificar se esses transcritos intrônicos possuíam orientação senso e/ou antissenso em relação aos seus respectivos mRNAs utilizando RT fita específica. Para isso alíquotas de RNA de pacientes tumorais (tanto tumor primário quanto metástase) foram utilizadas e mostraram que 4 transcritos intrônicos possuem orientação antissenso em relação ao mRNA codificador de proteína: *PPP3CB*, *ATF2*, *TGFBR2* e *MAPK1*. A transcrição antissenso para o transcrito *PPP3CB* também foi detectada na linhagem MIA PaCa-2. A orientação antissenso relativa ao correspondente codificador de proteína mostra uma forte evidência que suporta a hipótese de que este transcrito é uma unidade transcricional independente, que está sob o controle de um promotor diferente do codificador de proteína de mesmo *locus*.

Transcritos mapeando em regiões intrônicas com a mesma orientação ao mRNA codificador de proteína correspondente foram detectados nos *loci ATF2*, *TGFBR2* e *MAPK1*, assim como em 7 outros *loci* testados (*ARRB1*, *MAP3K1*, *MAP3K14*, *MAP2K5*, *PTEN*, *DAPK1* e *RAPGF2*), nos dois tipos de amostras de tecidos de paciente e amostras de RNA de MIA PaCa-2. Neste caso fica muito difícil de determinar se esses transcritos, que possuem orientação senso ao respectivo codificador de proteína, são unidades transcricionais independentes, pois também poderiam ser resultado da transcrição reversa de RNAs não

processados ou RNAs *lariats* estáveis provenientes do processamento de pré-mRNA. Experimentos serão necessários para identificar se de fato esses RNAs não codificadores de orientação senso são originados de unidades transcricionais independentes.

A abundância relativa de dois RNAs sensos (*DAPK1* e *MAP3K14*) e um antissenso (*PPP3CB*), mapeados em regiões intrônicas, foram medidos em amostras de tumores pancreáticos primários e metastáticos por qRT-PCR, confirmando os resultados obtidos pelas análises de microarranjos anteriormente. Outros 4 lncRNAs intrônicos mostraram resultados concordantes entre qRT-PCR e as hibridizações de microarranjos (*ARRB*, *RAPGF2*, *ATF2* e *PTEN*) apesar não serem estatisticamente significantes. Mudanças de expressão de 4 outros intrônicos não foram concordantes entre qRT-PCR e as hibridizações de microarranjos (*MAP3K1*, *TGFR2*, *MAP2K5* e *MAPK1*). A quantidade de RNA e o número de amostras de tecido de pacientes era muito limitado, assim os lncRNAs que não foram validados foram testados em apenas pequenas quantidades (n=3 para cada grupo) de amostras de pacientes em uma primeira etapa de validação. É possível que alguns desses candidatos de lncRNAs intrônicos que falharam na primeira etapa de validação, provavelmente, poderiam ser validados como diferencialmente expressos por qRT-PCR se mais amostras tivessem disponíveis. Entretanto, a não validação da diferença de expressão dos lncRNAs encontrados também poderia ser explicada pela presença de hibridação cruzada nas sondas da plataforma de microarranjos, ou pelo fato de haver algum tipo de viés na etapa de amplificação de RNA, que foi necessária para hibridação das amostras a plataforma de microarranjos, enquanto que para o experimento de qRT-PCR RNA total (sem etapas de amplificação) foi utilizado.

Alguns transcritos provenientes de regiões intragênicas já haviam sido demonstrados por modular em *cis* a expressão de mRNAs expressos no mesmo *locus* gênico como (*KDELR2* (Louro et al. 2007), *Saf* (Yan et al. 2005), *Zeb2-NAT* (Beltran et al. 2008)). No

trabalho de Louro e colaboradores (Louro et al. 2007), o aumento do transcrito antissenso *KDELR2* após tratamento com andrógeno era acompanhado com maior retenção do exon 2 e menor expressão do exon 1 do mRNA correspondente nesse *locus*, sugerindo que o RNA antissenso possa estar regulando o aparecimento de isoformas de *splicing* do correspondente codificador de proteína. A superexpressão do antissenso *Saf*, transcrito a partir do primeiro intron do *locus Fas*, aumentou a resistência à apoptose das células Jukart mediada por *Fas* mas não por *TNF-α*. A presença do RNA *Saf* não afeta a expressão total do *locus Fas*, mas sim o balanço das isoformas (Yan et al. 2005). Outro caso interessante é a expressão do transcrito antissenso *Zeb2-NAT* (*Zeb2-Natural Antisense Transcript*). Em células epiteliais, Zeb2-NAT não é transcrito, assim o *splicing* do codificador de proteína ocorre naturalmente, gerando um transcrito com 5'UTR de 421 nt. Este 5'UTR possui uma sequência que inibe o *scanning* dos ribossomos prevenindo a tradução da proteína Zeb2. Em células mesenquimais, o Zeb2-NAT é transcrito e previne a ligação do spliceossomo no primeiro intron. Assim, o transcrito processado possui a sequência do primeiro intron, que carrega a sequência IRES situada perto no início de tradução, permitindo que a proteína zeb2 seja traduzida (Beltran et al. 2008). Além disso, nosso grupo de pesquisa mostrou lncRNA capaz de regular a expressão em *cis* do seu respectivo codificador de proteína denominado *ANRASSF1*, que é responsável pela ligação do complexo de proteína *polycomb* (PRC2) para a cromatina da região promotora da isoforma *RASSF1A*, modulando sua expressão e a proliferação celular (Beckedorff, 2012)[21]. Desta forma, fomos investigar qual era o nível de expressão dos respectivos codificadores de proteínas correspondentes aos *loci PPP3CB*, *MAP3K14* e *DAPK1* nas mesmas amostras e não foi observada nenhuma diferença estatística entre as amostras tumorais e metastáticas. Este resultado indica que os transcritos intrônicos produzidos nesses

[21] Beckedorff, F.C.F. "Recrutamento do complexo repressivo polycomb 2 pelo RNA não codificador longo antissenso ANRASSF1 modula a expressão do gene RASSF1A e a proliferação celular." (2012). Tese (Doutorado em Bioquímica) – Instituto de Química, Universidade de São Paulo, São Paulo, 2012.

loci não parecem afetar em *cis* a abundância do transcrito codificador de proteína correspondente. Esta conclusão ainda é suportada pela falta de correlação significativa entre os níveis de expressão dos codificadores de proteína e os RNAs não codificadores originados dos *loci PPP3CB* e *DAPK1*. Devido à falta de correlação entre os níveis de expressão desses 2 lncRNAs e seus respectivos codificadores de proteína, levanta a hipótese de que esses lncRNAs possam estar agindo em *trans* ou então não possuem influencia no nível transcricional, mas no nível traducional de proteínas.

Para poder começar qualquer tipo de estudo funcional desses lncRNAs associados à metástase pancreática, seria interessante tentar determinar qual seria o tamanho e sequência desses lncRNAs. A primeira tentativa foi utilizar bibliotecas comerciais de CLONETECH de pâncreas e fazer RACE 5' e 3'. Nesta primeira abordagem, os lncRNAs proveninentes de ESTs de tamanho de cerca de 500pb foram identificados com tamanhos maiores, de no mínimo 1kb, sugerindo que esses RNAs não codificadores sejam longos. Com intuito de eliminar a possibilidade de se trabalhar com um transcrito intrônico originado de pre-mRNA ou produtos do processamento pós-transcricional, o transcrito antissenso foi escolhido para os experimentos de *northern blot* realizados. Muitas tentativas de caracterização do tamanho do lncRNA antissenso originado da região intrônica *PPP3CB* foram realizadas. As estratégias adotadas foram diferentes métodos de detecção (radioativo e não radioativo), uso de maior quantidade de RNA total ou RNA enriquecido com mRNA. Entretanto, os resultados obtidos até o momento foram inconclusivos a respeito do tamanho deste lncRNA (ficando entre 1,5-1,8kb). Entretanto, a técnica de *northern blotting* se mostrou pouco sensível para a detecção de transcritos expressos em baixo número de cópias como os ncRNAs intrônicos investigados neste trabalho. Outros grupos também já apresentaram dificuldades para identificar ncRNAs por meio de *Northern Blot*, como o trabalho de Braconi e colaboradores (Braconi et al. 2010) para identificar o transcrito não codificador TUC338, o qual apresentou uma alta taxa de

background, e o trabalho de Lu e colaboradores (Lu et al. 2010) investigou cinco ncRNAs testados, somente 3 foram detectados no *Northern Blot*. Os autores justificaram que a ocorrência deste fato foi devido à sensibilidade da técnica, que para alguns casos acaba sendo limitante.

Desta forma, a estratégia foi tentar confirmar se pelo menos o tamanho do transcrito antissenso apresentado pela biblioteca de RACE também se confirmaria em linhagem celular MIA PaCa-2. Infelizmente, não foi possível identificar o transcrito descrito pela biblioteca de RACE na linhagem MIA PaCa-2, o que impossibilitou a continuação de estudos relacionados ao lncRNA *PPP3CB*. Este resultado pode ser derivado do fato de que as bibliotecas de RACE são sintetizadas a partir de RNA total de tecidos pancreáticos normais e a linhagem MIA PaCa-2 é derivada de carcinoma primário. Isto pode ser reforçado com os resultados dos experimentos de RT-fita-específica (Figura 26), os quais mostram que 3 transcritos intrônicos são identificados em ambas orientações, enquanto que para a linhagem MIA PaCa-2, estes mesmos transcritos são observados apenas na orientação senso, sugerindo que para estes transcritos a linhagem celular pancreática não consegue mimetizar o que ocorre na situação real do tumor pancreático. Ainda, diferenças de expressão entre tecidos pancreáticos tumorais e linhagens celulares já foram encontradas no trabalho Lowe e colaboradores (Lowe et al. 2007), o que suporta os argumentos apresentados acima.

Devido ao fato do transcrito antissenso *PPP3CB* não ter alcançado resultados promissores, foi decido escolher outro candidato para caracterização. O próximo candidato escolhido foi o transcrito de orientação senso *DAPK1* por dois motivos interessantes. O primeiro seria a falta de correlação entre a expressão do transcrito intrônico com o seu respectivo codificador de proteína, sugerindo que este transcrito possa ser uma unidade transcricional independente, e, o segundo, é que a proteína Dapk1 possuía atividade pró-

apoptótica (Kogel et al. 2001) e trabalhos de imunohistoquímica em pacientes com tumor pancreático metastático identificou uma menor taxa da proteína quando comparados aos pacientes com tumores primários (Dansranjavin et al. 2006). Foi visto na literatura que a função pró apotótica desta proteína estava associada à presença do *"Death Domain"* (Stevens et al. 2009) e esse domínio se localiza na região C terminal. Interessantemente, isoformas mais curtas que não apresentavam o *Death-Domain* foram identificadas pelo no banco de dados de ESTs depositadas e visualizadas no *Genome Browser* e estavam associadas aos tecidos tumorais como mostrado na Figura 37. Estas evidências levantaram a hipótese que talvez o transcrito intrônico antisenso *DAPK1* pudesse modular o nível de expressão das duas isoformas. Entretanto, antes que esta hipótese fosse levada adiante, foi necessário verificar se o transcrito intrônico senso *DAPK-1* poderia representar uma unidade transcricional independente regulado por outro promotor que não fosse determinado pelo promotor do codificador de proteína. Assim, análises bioinformáticas foram utilizadas para a busca de uma possível região promotora. Os programas utilizados apontaram por uma possível região promotora e dois sítios de ligação para a proteína TBP (*TATA-bindig protein*) e ainda a evidência foi reforçada pela presença de marcas de H3K4me3, a qual está associada a presença de região promotora ativa (Guttman et al. ; Mikkelsen et al. 2007), em linhagem celular não pancreática GM13878 (Figura 5).

O ensaio de atividade promotora com gene repórter da luciferase permitiu identificar uma região promotora forte com 50% da atividade promotora do vetor pSV40. Assim, deu-se continuidade aos estudos relacionados ao lncRNA intrônico *DAPK-1*. Como já havia observado anteriormente, o *locus* referente ao gene *DAPK-1* possui uma isoforma mais curta que foi identificada em nas ESTs *spliced* provenientes de tecidos tumorais de língua e pele. Assim, como já descrito por outro trabalho (Yan et al. 2005), o RNA não codificador intrônico antissenso possui a capacidade de modular isoformas de *splicing* sem alterar

substancialmente a expressão total do *locus* do gene *Fas*. Uma vez que os resultados referentes à expressão de todas as isoformas codificadoras de dapk-1 (mRNAs) não estava correlacionada com a expressão do intrônico *DAPK-1*, procurou-se observar se este transcrito intrônico, apesar de ter orientação senso, poderia estar correlacionado com a presença da isoforma longa nas amostras clínicas de pacientes. Com os resutlados obtidos das medidas de expressão por qPRC, não foi possível ver associação do nível de expressão de mRNA da isoforma codificadora longa com o nível de expressão do lncRNA intrônico *DAPK-1* quando analisada nos tecidos pancreáticos de amostras clínicas de pacientes tumorais. Entretanto, curiosamente, utilizando as linhagens pancreáticas (MIA PaCa-2 e Capan-1) foi possível verificar que os níveis de expressão entre esses dois RNAs estudados possuíam uma alta correlação de expressão. Isto pode ser explicado, pelo fato, das linhagens pancreáticas apresentarem apenas uma parte isolada do tumor, enquanto que as amostras clínicas tumorais possuíam uma população heterogênea de células, incluindo células estromais, que, na maioria das vezes acaba sendo grande parte da amostra retirada do paciente (Erkan et al. 2012; Kong et al. 2012).

Devido à alta correlação na expressão do ncRNA intrônico e da isoforma longa do *locus* DAPK1em linhagens celulares (MIA PaCa-2 e Capan-2), fomos verificar se isso se era possível discriminar entre a expressão do lncRNA intrônico *DAPK-1* e o respectivo pré-mRNA. Como esperado, a expressão do não codificador se apresentou mais alta do que o seu respectivo pré-mRNA. Desta forma, fomos verificar se o silenciamento do lncRNA intrônico provocaria uma mudança na expressão da isoforma mais longa. Para isso, a linhagem Capan-1 foi escolhida por apresentar maior expressão do transcrito intrônico. Entretanto, os resultados de silenciamento, mostraram que o oligo antissenso modificado utilizado era capaz de silenciar tanto o lncRNA quanto as outras isoformas codificadoras. Para checar se o efeito do silenciamento das isoformas codificadores não eram efeito secundário do silenciamento do

lncRNA, foram desenhados 3 pares *primers* em regiões intrônicas sem evidência de transcrição para medir a quantidade de pré-mRNA do *locus DAPK-1*. Como mostrado na Figura 42, o ensaio de silenciamento utilizando oligo antissenso estava diminuindo a expressão do pré-mRNA, o que poderia ter causado a diminuição do mRNA codificador de proteína na Figura 41. Entretanto, esse experimento, porém, não pode descartar a possibilidade de o silenciamento do ncRNA *DAPK-1* estar envolvido com algum mecanismo que iniba a transcrição no *locus* de *DAPK-1*, e a diminuição na expressão do pré-mRNA seja um efeito secundário da ausência do ncRNA *DAPK-1*.

Apesar dos resultados referentes ao possível papel do intrônico senso *DAPK-1* ainda serem muito iniciais, na literatura ainda há poucos trabalhos que exploram a possibilidade de funcionalidade de transcritos intrônicos senso. Entretanto diversos trabalhos na literatura relatam a presença de transcritos provenientes de regiões intrônicas, o término do projeto ENCODE identificou que 62% das bases genômicas são representadas por RNAs longos (>200 nucleotídeos). Dessas bases apenas 5.5% são representadas por exons da anotação do GENENCODE, outros 94.5% são, na maior parte, representadas por regiões intrônicas, sendo que apenas 31% são provenientes de regiões intergênicas (Consortium et al. 2012). Levando em consideração os transcritos primários que ainda não foram processados a porcentagem que é transcrita no genoma sobe para 74.7% (Djebali et al. 2012). Kapranov e colaboradores (Kapranov et al. 2010) demonstraram que a maior parte da porção não poliadenilada do RNA, excluindo a fração ribossomal, está mapeada em regiões intrônicas com taxas variando de 31%-47% em humanos (K562- leucemia, fígado e cérebro), mostrando que, de fato, RNAs provenientes de regiões intrônicas representam uma fração considerável de RNAs não poliadenilados. No trabalho de St Laurent e colaboradores (St Laurent et al. 2012), RNA-seq de 42 ratos tratados com LPS (lipopolissacarídeo) mostraram que 5753 regiões intrônicas apresentaram densidade de reads maior do que a dos exons do correspondente *locus* em pelo

menos um dos animais tratados. Destes 707 apresentavam razão de densidade de *reads* (intron/exon) de 2 a 10 vezes maior do que os de região exônica dos respectivos *locus*, sendo que 90% apresentavam transcritos intrônicos não anotados, outros 10% representavam snoRNAs e regiões repetitivas do genoma, sugerindo que estes transcritos intrônicos sejam funcionais.

Apesar das evidências de que a transcrição no genoma é pervasiva (Carninci et al. 2005; Dinger et al. 2009; Kapranov and St Laurent 2012), pouca atenção foi dada para RNAs intrônicos senso, devido ao fato da dificuldade de diferenciar este transcrito como uma unidade transcricional independente, ou produtos resultantes da transcrição espúria da RNA polimerase II em regiões que se encontram acessíveis para permitir a expressão de genes codificadores de proteína (Struhl 2007; van Bakel et al. 2010), ou serem pré-mRNAs de genes abundantes, ou ainda, serem produtos do processamento do *splicing* de pre-mRNAs sendo denominados intron *lariats*. Entretanto, há na literatura trabalhos que propõem funcionalidade para os intrônicos senso muito parecidas com funções descritas para lncRNAs intergênicos e antissenso.

O transcrito *COLDAIR* (*COLD Assisted Intronic Noncondig RNA*) de cerca de 1100 nt foi identificado em *Arabidopsis thaliana* e possui orientação senso em relação ao mRNA do gene *FLC* (*Flowering Locus C*). Este transcrito foi identificado como sendo capaz de direcionar o complexo PCR2 para região promotora do seu gene hospedeiro (*FLC*), causando a repressão deste transcrito no processo de vernalização (Heo and Sung 2011). Este processo de regulação por meio do recrutamento do complexo PCR2 já foi descrito em animais para o lncRNA HOTAIR, transcritos a partir da família de genes *HOXC* e recruta o complexo proteico *polycomb* para silenciar em *trans* o *locus HOXD* (Rinn et al. 2007; Tsai et al. 2010). A superexpressão de 3 regiões intrônicas (6a, 14b e 23) do gene *CFTR* (*cystic fibrosis*

transmembrane conductance regulator) em linhagem celular HeLa, a qual não expressa o codificador de proteína, foi capaz de afetar genes relacionados com *CFTR* (Hill et al. 2006), mostrando que regiões intrônicas senso possuem papéis importantes na célula. Recentemente, trabalho realizado por Khaitan e colaboradores descreveu um transcrito intrônico senso *SPY4-IT1*, o qual era transcrito a paritr do segundo intron do gene *SPYR4*, em melanoma. Este transcrito intrônico apesar de possuir alta correlação de expressão com as isoformas codificadoras *SPRY4.2* e *SPRY4.4* parece não atuar em *cis*. O silenciamento deste transcrito levou a menor viabiliadade celular e aumento da taxa de apoptose dose dependente, além de inibir invasão celular em uma taxa de 60-80% (Khaitan et al. 2011). Guil e colaboradores (Guil et al. 2012) identificaram diversos transcritos mapeados em regiões intrônicas de genes reguladores da cromatina ligados à proteína EZH2 (componente do complexo PCR2).

Como trabalhos na literatura mostravam que transcritos intrônicos senso exercem funções parecidas com lncRNAs conhecidos, as estratégias utilizadas para caraterização funcional do transcrito senso *DAPK-1* foram voltadas para funções de lncRNAs conhecidos. Por exemplo, a regulação do *splicing* no balanço de isoformas realizada pelo transcrito antissenso *saf* (Yan et al. 2005). Como os resultados ainda são pouco conclusivos sobre o possível papel do lncRNA intrônico senso *DAPK-1*, experimentos referentes à caracterização do tamanho total deste transcrito devem ser realizados para ensaio de superexpressão em linhagens celulares, uma vez que o silenciamento deste transcrito não produziu resultados satisfatórios. A superexpressão deste transcrito poderia indicar se este está associado ao balanço das isoformas codificadoras, agindo em *cis*, ou se o seu possível papel estaria associado a outros *loci* gênicos, agindo em *trans*.

6 Conclusões

Neste trabalho, foi observado que RNAs não codificadores originados de regiões intrônicas e intergênicas são expressos em tecidos tumorais e não tumorais. Análises *in silico* demonstraram que parte desses ncRNAs possuem em média tamanho 779nt e a maior parte não possuem potencial codificador, caracterizando-os como lncRNAs. A identificação desses lncRNAs em outros tecidos com dados de RNA-seq na literatura, confirma a existência desses RNAs, que podem apresentar um padrão de tecido-especificidade ou expressão ampla.

Enriquecimento de marcas de cromatina associadas às regiões promotoras, sugere que estes transcritos possam ser regulados. Apesar do pequeno conjunto de lncRNAs analisados, evidências *in silico* apontam para o fato da maior parte destes transcritos não sofrerem regulação de ilhas CpGs canônicas. A presença de CAGE*tags* sugere que estes transcritos sofram processamento pós-transcricional, e a falta de correlação entre a expressão de lncRNAs e seus respectivos codificadores de proteína e a identificação de transcritos antissenso, reforça a ideia de que estes transcritos possam ser transcritos independentes, em vez de pré-mRNAs ou intron *lariats.* Evidências de conservação de sequência, estrutura e expressão apontam para o fato de que estes transcritos possam possuir funcionalidade. Análises adicionais devem ser realizadas para determinar a biogênese desses RNAs.

Interessantemente, foram encontradas duas assinaturas de expressão que correlacionvam com o tumores primários e a metástase pancreática, as quais eram compostas por RNAs codificadores de proteína juntamente com lncRNAs de regiões intrônicas e intergênicas. Apesar da validação de poucos candidatos, os 3 candidatos intrônicos tiveram a expressão confirmada em amostras de tumores metastáticos em relação às amostras de tumor primário, sendo candidatos promissores para estudos de caracterização. Os experimentos de RACE apontaram que estes 3 candidatos apresentavam mais de 1 kb, caracterizando-os como

realmente transcritos longos. A confirmação do tamanho do transcrito antissenso *PPP3CB* não mostrou resultados promissores, melhorias na metodologia ou outras estratégias devem ser utilizadas para determinação do possível tamanho do transcrito.

Os experimentos de atividade promotora à montante do transcrito *DAPK1* aponta para o fato deste transcrito ser uma unidade transcricional independente, sendo um transcrito longo de pelo menos 1050bp. Resultados iniciais apontam para o fato da expressão deste lncRNA estar correlacionado com a expressão da isoforma longa em linhagens pancreáticas (MIA PaCa2 e Capan-1). Infelizmente, o silenciamento desse transcrito não foi propício para investigar a correlação entre lncRNA e ao seu respectivo codificador de proteína, devido ao silenciamento do pré-mRNA.

Estudos adicionais deverão ser realizados a fim de investigar a biogênese desses RNAs, assim como o seu papel funcional na célula a fim de entender a relevância biológica desses lncRNAs no adenocarcinoma pancreático para compreender melhor as bases moleculares dessa doença.

7 Referências

Abrahams, B. S., D. Tentler, et al. (2007). "Genome-wide analyses of human perisylvian cerebral cortical patterning." Proceedings of the National Academy of Sciences of the United States of America **104**(45): 17849-17854.

Adachi, J., C. Kumar, et al. (2006). "The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins." Genome Biol **7**(9): R80.

Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of Drosophila melanogaster." Science **287**(5461): 2185-2195.

Aloysius, M. M., A. M. Zaitoun, et al. (2010). "Mucins and CD56 as markers of tumour invasion and prognosis in periampullary cancer." Br J Surg **97**(8): 1269-1278.

Alter, O. and G. H. Golub (2006). "Singular value decomposition of genome-scale mRNA lengths distribution reveals asymmetry in RNA gel electrophoresis band broadening." Proc Natl Acad Sci U S A **103**(32): 11828-11833.

Anderson, N. L., M. Polanski, et al. (2004). "The human plasma proteome: a nonredundant list developed by combination of four separate sources." Mol Cell Proteomics **3**(4): 311-326.

Arumugam, T., V. Ramachandran, et al. (2009). "Epithelial to mesenchymal transition contributes to drug resistance in pancreatic cancer." Cancer Res **69**(14): 5820-5828.

Badea, L., V. Herlea, et al. (2008). "Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia." Hepatogastroenterology **55**(88): 2016-2027.

Baginsky, S., L. Hennig, et al. (2010). "Gene expression analysis, proteomics, and network discovery." Plant Physiol **152**(2): 402-410.

Baker, M. (2011). "Long noncoding RNAs: the search for function." Nat Methods **8**(5): 379-383.

Banfai, B., H. Jia, et al. (2012). "Long noncoding RNAs are rarely translated in two human cell lines." Genome Res **22**(9): 1646-1657.

Baniwal, S. K., O. Khalid, et al. "Runx2 transcriptome of prostate cancer cells: insights into invasiveness and bone metastasis." Mol Cancer **9**: 258.

Barski, A., S. Cuddapah, et al. (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-837.

Bartel, D. P. (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." Cell **116**(2): 281-297.

Bejerano, G., M. Pheasant, et al. (2004). "Ultraconserved elements in the human genome." Science **304**(5675): 1321-1325.

Beltran, M., I. Puig, et al. (2008). "A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition." Genes Dev **22**(6): 756-769.

Belyaev, O., M. H. Seelig, et al. (2008). "Intraductal papillary mucinous neoplasms of the pancreas." J Clin Gastroenterol **42**(3): 284-294.

Benito, M., J. Parker, et al. (2004). "Adjustment of systematic microarray data biases." Bioinformatics **20**(1): 105-114.

Bertone, P., V. Stolc, et al. (2004). "Global identification of human transcribed sequences with genome tiling arrays." Science **306**(5705): 2242-2246.

Biankin, A. V., N. Waddell, et al. (2012). "Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes." Nature **491**(7424): 399-405.

Birney, E., J. A. Stamatoyannopoulos, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Blackhall, F. H., M. Pintilie, et al. (2004). "Stability and heterogeneity of expression profiles in lung cancer specimens harvested following surgical resection." Neoplasia **6**(6): 761-767.

Bloomston, M., W. L. Frankel, et al. (2007). "MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis." JAMA **297**(17): 1901-1908.

Board, N. E. (2000). "The nature of the number." Nat Genet **25**(2): 127-128.

Bolstad, B. M., R. A. Irizarry, et al. (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." Bioinformatics **19**(2): 185-193.

Braconi, C., N. Valeri, et al. (2010). "Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma." Proc Natl Acad Sci U S A **108**(2): 786-791.

Branham, W. S., C. D. Melvin, et al. (2007). "Elimination of laboratory ozone leads to a dramatic improvement in the reproducibility of microarray gene expression measurements." BMC Biotechnol **7**: 8.

Brito, G. C., A. A. Fachel, et al. (2008). "Identification of protein-coding and intronic noncoding RNAs down-regulated in clear cell renal carcinoma." Mol Carcinog **47**(10): 757-767.

Brown, C. J., A. Ballabio, et al. (1991). "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome." Nature **349**(6304): 38-44.

Buchholz, M., W. Boeck, et al. (2001). "Use of DNA arrays/microarrays in pancreatic research." Pancreatology **1**(6): 581-586.

Buchholz, M., M. Braun, et al. (2005). "Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions." Oncogene **24**(44): 6626-6636.

Buchholz, M., H. Kestler, et al. (2008). "Differential diagnosis of pancreatic tumors by molecular analysis of clinical specimens." Pancreatology **8**(6): 551-557.

Cabili, M. N., C. Trapnell, et al. (2011). "Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses." Genes Dev **25**(18): 1915-1927.

Caldas, C., S. A. Hahn, et al. (1994). "Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma." Nat Genet **8**(1): 27-32.

Calin, G. A., C. G. Liu, et al. (2007). "Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas." Cancer Cell **12**(3): 215-229.

Camargo, A. A., H. P. Samaia, et al. (2001). "The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome." Proc Natl Acad Sci U S A **98**(21): 12103-12108.

Campagna, D., L. Cope, et al. (2008). "Gene expression profiles associated with advanced pancreatic cancer." Int J Clin Exp Pathol **1**(1): 32-43.

Campbell, P. J., S. Yachida, et al. (2010). "The patterns and dynamics of genomic instability in metastatic pancreatic cancer." Nature **467**(7319): 1109-1113.

Candeias, M. M., L. Malbert-Colas, et al. (2008). "P53 mRNA controls p53 activity by managing Mdm2 functions." Nat Cell Biol **10**(9): 1098-1105.

Cao, D., R. Ashfaq, et al. (2008). "Differential Expression of Multiple Genes in Association with MADH4/DPC4/SMAD4 Inactivation in Pancreatic Cancer." Int J Clin Exp Pathol **1**(6): 510-517.

Capurso, G., S. Lattimore, et al. (2006). "Gene expression profiles of progressive pancreatic endocrine tumours and their liver metastases reveal potential novel markers and therapeutic targets." Endocr Relat Cancer **13**(2): 541-558.

Carninci, P., T. Kasukawa, et al. (2005). "The transcriptional landscape of the mammalian genome." Science **309**(5740): 1559-1563.

Cawley, S., S. Bekiranov, et al. (2004). "Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs." Cell **116**(4): 499-509.

Chelala, C., S. A. Hahn, et al. (2007). "Pancreatic Expression database: a generic model for the organization, integration and mining of complex cancer datasets." BMC Genomics **8**: 439.

Chen, G., K. Yin, et al. (2011). "Comparative analysis of human protein-coding and noncoding RNAs between brain and 10 mixed cell lines by RNA-Seq." PLoS One **6**(11): e28318.

Chen, J., M. Sun, et al. (2002). "Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags." Proc Natl Acad Sci U S A **99**(19): 12257-12262.

Chen, R., T. A. Brentnall, et al. (2007). "Quantitative proteomics analysis reveals that proteins differentially expressed in chronic pancreatitis are also frequently involved in pancreatic cancer." Mol Cell Proteomics **6**(8): 1331-1342.

Chen, R., S. Pan, et al. (2006). "Quantitative proteomic profiling of pancreatic cancer juice." Proteomics **6**(13): 3871-3879.

Chen, R., E. C. Yi, et al. (2005). "Pancreatic cancer proteome: the proteins that underlie invasion, metastasis, and immunologic escape." Gastroenterology **129**(4): 1187-1197.

Cheng, J., P. Kapranov, et al. (2005). "Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution." Science **308**(5725): 1149-1154.

Cheng, Z. and T. M. Menees (2011). "RNA splicing and debranching viewed through analysis of RNA lariats." Mol Genet Genomics **286**(5-6): 395-410.

Clark, M. B. and J. S. Mattick (2011). "Long noncoding RNAs in cell biology." Semin Cell Dev Biol.

Conroy, T., F. Desseigne, et al. (2011). "FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer." N Engl J Med **364**(19): 1817-1825.

Consortium, C. e. S. (1998). "Genome sequence of the nematode C. elegans: a platform for investigating biology." Science **282**(5396): 2012-2018.

Consortium, E. P., I. Dunham, et al. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Consortium, I. H. G. S. (2004). "Finishing the euchromatic sequence of the human genome." Nature **431**(7011): 931-945.

Coram, T. E., M. L. Settles, et al. (2008). "Transcriptome analysis of high-temperature adult-plant resistance conditioned by Yr39 during the wheat-Puccinia striiformis f. sp tritici interaction." Molecular Plant Pathology **9**(4): 479-493.

Creyghton, M. P., A. W. Cheng, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Crick, F. (1970). "Central dogma of molecular biology." Nature **227**(5258): 561-563.

Crick, F. H. (1958). "On protein synthesis." Symp Soc Exp Biol **12**: 138-163.

Crippa, S., C. Fernandez-Del Castillo, et al. (2010). "Mucin-producing neoplasms of the pancreas: an analysis of distinguishing clinical and epidemiologic characteristics." Clin Gastroenterol Hepatol **8**(2): 213-219.

Crnogorac-Jurcevic, T., E. Efthimiou, et al. (2001). "Gene expression profiles of pancreatic cancer and stromal desmoplasia." Oncogene **20**(50): 7437-7446.

Crnogorac-Jurcevic, T., E. Efthimiou, et al. (2002). "Expression profiling of microdissected pancreatic adenocarcinomas." Oncogene **21**(29): 4587-4594.

Crnogorac-Jurcevic, T., R. Gangeswaran, et al. (2005). "Proteomic analysis of chronic pancreatitis and pancreatic adenocarcinoma." Gastroenterology **129**(5): 1454-1463.

Crnogorac-Jurcevic, T., E. Missiaglia, et al. (2003). "Molecular alterations in pancreatic carcinoma: expression profiling shows that dysregulated expression of S100 genes is highly prevalent." J Pathol **201**(1): 63-74.

Cutts, R. J., E. Gadaleta, et al. (2010). "The Pancreatic Expression database: 2011 update." Nucleic Acids Res **39**(Database issue): D1023-1028.

Dansranjavin, T., C. Mobius, et al. (2006). "E-cadherin and DAP kinase in pancreatic adenocarcinoma and corresponding lymph node metastases." Oncol Rep **15**(5): 1125-1131.

Deng, H., J. Shi, et al. (2008). "Usefulness of S100P in diagnosis of adenocarcinoma of pancreas on fine-needle aspiration biopsy specimens." Am J Clin Pathol **129**(1): 81-88.

Dennis, G., Jr., B. T. Sherman, et al. (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." Genome Biol **4**(5): P3.

Dias Neto, E., R. G. Correa, et al. (2000). "Shotgun sequencing of the human transcriptome with ORF expressed sequence tags." Proc Natl Acad Sci U S A **97**(7): 3491-3496.

DiMagno, E. P., H. A. Reber, et al. (1999). "AGA technical review on the epidemiology, diagnosis, and treatment of pancreatic ductal adenocarcinoma. American Gastroenterological Association." Gastroenterology **117**(6): 1464-1484.

Ding, L. C., L. She, et al. "Notch-4 contributes to the metastasis of salivary adenoid cystic carcinoma." Oncol Rep **24**(2): 363-368.

Dingemans, A.-M., B. Vosse, et al. (2008). "Profiling of integrin (ITG) expression in early stage non-small cell lung cancer (NSCLC)." AACR Meeting Abstracts **2008**(1_Annual_Meeting): 2161-.

Dinger, M. E., P. P. Amaral, et al. (2009). "Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications." Brief Funct Genomic Proteomic.

Dinger, M. E., K. C. Pang, et al. (2008). "Differentiating protein-coding and noncoding RNA: challenges and ambiguities." PLoS Comput Biol **4**(11): e1000176.

Dite, P., M. Hermanova, et al. (2012). "The role of chronic inflammation: chronic pancreatitis as a risk factor of pancreatic cancer." Dig Dis **30**(3): 277-283.

Djebali, S., C. A. Davis, et al. (2012). "Landscape of transcription in human cells." Nature **489**(7414): 101-108.

Dowen, S. E., T. Crnogorac-Jurcevic, et al. (2005). "Expression of S100P and its novel binding partner S100PBPR in early pancreatic cancer." Am J Pathol **166**(1): 81-92.

Du, T. and P. D. Zamore (2005). "microPrimer: the biogenesis and function of microRNA." Development **132**(21): 4645-4652.

Dudoit, S., M. J. van der Laan, et al. (2004). "Multiple testing. Part I. Single-step procedures for control of general type I error rates." Stat Appl Genet Mol Biol **3**: Article13.

Dunham, I., A. Kundaje, et al. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Dupuy, A. and R. M. Simon (2007). "Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting." J Natl Cancer Inst **99**(2): 147-157.

Ebert, M. S. and P. A. Sharp (2012). "Roles for microRNAs in conferring robustness to biological processes." Cell **149**(3): 515-524.

Eberwine, J. (1996). "Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA." Biotechniques **20**(4): 584-591.

Erdmann, V. A., M. Z. Barciszewska, et al. (2001). "Regulatory RNAs." Cell Mol Life Sci **58**(7): 960-977.

Erkan, M., S. Hausmann, et al. (2012). "The role of stroma in pancreatic cancer: diagnostic and therapeutic implications." Nat Rev Gastroenterol Hepatol **9**(8): 454-467.

Ewing, B. and P. Green (2000). "Analysis of expressed sequence tags indicates 35,000 human genes." Nat Genet **25**(2): 232-234.

Faghihi, M. A. and C. Wahlestedt (2009). "Regulatory roles of natural antisense transcripts." Nat Rev Mol Cell Biol **10**(9): 637-643.

Fare, T. L., E. M. Coffey, et al. (2003). "Effects of atmospheric ozone on microarray data quality." Anal Chem **75**(17): 4672-4675.

Feig, C., A. Gopinathan, et al. (2012). "The pancreas cancer microenvironment." Clin Cancer Res **18**(16): 4266-4276.

Fejes-Toth, K., V. Sotirova, et al. (2009). "Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs." Nature **457**(7232): 1028-1032.

Fong, Z. V. and J. M. Winter (2012). "Biomarkers in pancreatic cancer: diagnostic, prognostic, and predictive." Cancer J **18**(6): 530-538.

Forrest, A. R., R. F. Abdelhamid, et al. (2009). "Annotating non-coding transcription using functional genomics strategies." Brief Funct Genomic Proteomic **8**(6): 437-443.

Friess, H., J. Ding, et al. (2003). "Microarray-based identification of differentially expressed growth- and metastasis-associated genes in pancreatic cancer." Cell Mol Life Sci **60**(6): 1180-1199.

Frith, M. C., A. R. Forrest, et al. (2006). "The abundance of short proteins in the mammalian proteome." PLoS Genet **2**(4): e52.

Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-282.

Gibb, E. A., C. J. Brown, et al. (2011). "The functional role of long non-coding RNA in human carcinomas." Mol Cancer **10**: 38.

Gibb, E. A., E. A. Vucic, et al. (2011). "Human cancer long non-coding RNA transcriptomes." PLoS One **6**(10): e25915.

Gironella, M., M. Seux, et al. (2007). "Tumor protein 53-induced nuclear protein 1 expression is repressed by miR-155, and its restoration inhibits pancreatic tumor development." Proc Natl Acad Sci U S A **104**(41): 16170-16175.

Gnjatic, S., E. Ritter, et al. (2010). "Seromic profiling of ovarian and pancreatic cancer." Proc Natl Acad Sci U S A **107**(11): 5088-5093.

Green, P. and B. Ewing. (2002). from http://www.phrap.org.

Greer, J. B., D. C. Whitcomb, et al. (2007). "Genetic predisposition to pancreatic cancer: a brief review." Am J Gastroenterol **102**(11): 2564-2569.

Gress, T. M., F. Muller-Pillasch, et al. (1996). "A pancreatic cancer-specific expression profile." Oncogene **13**(8): 1819-1830.

Gress, T. M., C. Wallrapp, et al. (1997). "Identification of genes with specific expression in pancreatic cancer by cDNA representational difference analysis." Genes Chromosomes Cancer **19**(2): 97-103.

Griffiths-Jones, S., H. K. Saini, et al. (2008). "miRBase: tools for microRNA genomics." Nucleic Acids Res **36**(Database issue): D154-158.

Grivennikov, S. I., F. R. Greten, et al. (2010). "Immunity, inflammation, and cancer." Cell **140**(6): 883-899.

Gronborg, M., J. Bunkenborg, et al. (2004). "Comprehensive proteomic analysis of human pancreatic juice." J Proteome Res **3**(5): 1042-1055.

Gruber, A. R., R. Neubock, et al. (2007). "The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures." Nucleic Acids Res **35**(Web Server issue): W335-338.

Grutzmann, R., H. Boriss, et al. (2005). "Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes." Oncogene **24**(32): 5079-5088.

Grutzmann, R., C. Pilarsky, et al. (2004). "Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays." Neoplasia **6**(5): 611-622.

Guerra, C., A. J. Schuhmacher, et al. (2007). "Chronic pancreatitis is essential for induction of pancreatic ductal adenocarcinoma by K-Ras oncogenes in adult mice." Cancer Cell **11**(3): 291-302.

Guil, S., M. Soler, et al. (2012). "Intronic RNAs mediate EZH2 regulation of epigenetic targets." Nat Struct Mol Biol **19**(7): 664-670.

Guo, S., M. A. Hakimi, et al. (2005). "Linking transcriptional elongation and messenger RNA export to metastatic breast cancers." Cancer Res **65**(8): 3011-3016.

Gupta, R. A., N. Shah, et al. (2010). "Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis." Nature **464**(7291): 1071-1076.

Guttman, M., I. Amit, et al. (2009). "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals." Nature **458**(7235): 223-227.

Guttman, M., M. Garber, et al. "Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs." Nat Biotechnol **28**(5): 503-510.

Hahn, S. A., M. Schutte, et al. (1996). "DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1." Science **271**(5247): 350-353.

Hamada, S. and T. Shimosegawa (2011). "Biomarkers of pancreatic cancer." Pancreatology **11 Suppl 2**: 14-19.

Harrison, P. M., A. Kumar, et al. (2002). "A question of size: the eukaryotic proteome and the problems in defining it." Nucleic Acids Res **30**(5): 1083-1090.

Harrow, J., A. Frankish, et al. (2012). "GENCODE: the reference human genome annotation for The ENCODE Project." Genome Res **22**(9): 1760-1774.

Harsha, H. C., K. Kandasamy, et al. (2009). "A compendium of potential biomarkers of pancreatic cancer." PLoS Med **6**(4): e1000046.

Hashimoto, K., E. Ishida, et al. (2009). "A liver X receptor (LXR)-beta alternative splicing variant (LXRBSV) acts as an RNA co-activator of LXR-beta." Biochem Biophys Res Commun **390**(4): 1260-1265.

He, S., S. Liu, et al. (2011). "The sequence, structure and evolutionary features of HOTAIR in mammals." BMC Evol Biol **11**: 102.

Heo, J. B. and S. Sung (2011). "Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA." Science **331**(6013): 76-79.

Hezel, A. F., A. C. Kimmelman, et al. (2006). "Genetics and biology of pancreatic ductal adenocarcinoma." Genes Dev **20**(10): 1218-1249.

Hill, A. E., J. S. Hong, et al. (2006). "Micro-RNA-like effects of complete intronic sequences." Front Biosci **11**: 1998-2006.

Hofacker, I. L., M. Fekete, et al. (2002). "Secondary structure prediction for aligned RNA sequences." J Mol Biol **319**(5): 1059-1066.

Hosack, D. A., G. Dennis, Jr., et al. (2003). "Identifying biological themes within lists of genes with EASE." Genome Biol **4**(10): R70.

Hruban, R. H., M. Goggins, et al. (2000). "Progression model for pancreatic cancer." Clin Cancer Res **6**(8): 2969-2972.

Hu, L., S. Evers, et al. (2004). "Two-dimensional protein database of human pancreas." Electrophoresis **25**(3): 512-518.

Huang da, W., B. T. Sherman, et al. (2007). "DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists." Nucleic Acids Res **35**(Web Server issue): W169-175.

Huang, L., M. Zheng, et al. "Identification of a gene-expression signature for predicting lymph node metastasis in patients with early stage cervical carcinoma." Cancer.

Huarte, M. and J. L. Rinn (2010). "Large non-coding RNAs: missing links in cancer?" Hum Mol Genet **19**(R2): R152-161.

Hung, T. and H. Y. Chang (2012). "Long noncoding RNA in genome regulation: prospects and mechanisms." RNA Biol **7**(5): 582-585.

Iacobuzio-Donahue, C. A. (2012). "Genetic evolution of pancreatic cancer: lessons learnt from the pancreatic cancer genome sequencing project." Gut **61**(7): 1085-1094.

Iacobuzio-Donahue, C. A., A. Maitra, et al. (2003). "Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays." Am J Pathol **162**(4): 1151-1162.

Iacobuzio-Donahue, C. A., A. Maitra, et al. (2002). "Discovery of novel tumor markers of pancreatic cancer using global gene expression technology." Am J Pathol **160**(4): 1239-1249.

Iacobuzio-Donahue, C. A., R. E. Wilentz, et al. (2000). "Dpc4 protein in mucinous cystic neoplasms of the pancreas: frequent loss of expression in invasive carcinomas suggests a role in genetic progression." Am J Surg Pathol **24**(11): 1544-1548.

Initiative, T. A. G. (2000). "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana." Nature **408**(6814): 796-815.

Jarvinen, A. K., S. Hautaniemi, et al. (2004). "Are data from different gene expression microarray platforms comparable?" Genomics **83**(6): 1164-1168.

Jeffery, I. B., D. G. Higgins, et al. (2006). "Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data." BMC Bioinformatics **7**: 359.

Ji, P., S. Diederichs, et al. (2003). "MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer." Oncogene **22**(39): 8031-8041.

Jimenez, R. E., A. L. Warshaw, et al. (1999). "Sequential accumulation of K-ras mutations and p53 overexpression in the progression of pancreatic mucinous cystic neoplasms to malignancy." Ann Surg **230**(4): 501-509; discussion 509-511.

Johnson, W. E., C. Li, et al. (2007). "Adjusting batch effects in microarray expression data using empirical Bayes methods." Biostatistics **8**(1): 118-127.

Jones, S., X. Zhang, et al. (2008). "Core signaling pathways in human pancreatic cancers revealed by global genomic analyses." Science **321**(5897): 1801-1806.

Jongeneel, C. V., M. Delorenzi, et al. (2005). "An atlas of human gene expression from massively parallel signature sequencing (MPSS)." Genome Res **15**(7): 1007-1014.

Kampa, D., J. Cheng, et al. (2004). "Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22." Genome Res **14**(3): 331-342.

Kapranov, P., S. E. Cawley, et al. (2002). "Large-scale transcriptional activity in chromosomes 21 and 22." Science **296**(5569): 916-919.

Kapranov, P., J. Cheng, et al. (2007). "RNA maps reveal new RNA classes and a possible function for pervasive transcription." Science **316**(5830): 1484-1488.

Kapranov, P., J. Drenkow, et al. (2005). "Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays." Genome Res **15**(7): 987-997.

Kapranov, P. and G. St Laurent (2012). "Dark Matter RNA: Existence, Function, and Controversy." Front Genet **3**: 60.

Kapranov, P., G. St Laurent, et al. (2010). "The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA." BMC Biol **8**: 149.

Katayama, S., Y. Tomaru, et al. (2005). "Antisense transcription in the mammalian transcriptome." Science **309**(5740): 1564-1566.

Kavanaugh, L. A. and F. S. Dietrich (2009). "Non-coding RNA prediction and verification in Saccharomyces cerevisiae." PLoS Genet **5**(1): e1000321.

Khaitan, D., M. E. Dinger, et al. (2011). "The melanoma-upregulated long noncoding RNA SPRY4-IT1 modulates apoptosis and invasion." Cancer Res **71**(11): 3852-3862.

Khalil, A. M., M. Guttman, et al. (2009). "Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression." Proc Natl Acad Sci U S A **106**(28): 11667-11672.

Khan, J., L. H. Saal, et al. (1999). "Expression profiling in cancer using cDNA microarrays." Electrophoresis **20**(2): 223-229.

Kim, K., I. Jutooru, et al. (2012). "HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer." Oncogene.

Kim, W. Y. and N. E. Sharpless (2006). "The regulation of INK4/ARF in cancer and aging." Cell **127**(2): 265-275.

Kimura, W., A. Kuroda, et al. (1998). "Problems in the diagnosis and treatment of a so-called mucin-producing tumor of the pancreas." Pancreas **16**(3): 363-369.

Kindler, H. L., D. Niedzwiecki, et al. (2010). "Gemcitabine plus bevacizumab compared with gemcitabine plus placebo in patients with advanced pancreatic cancer: phase III trial of the Cancer and Leukemia Group B (CALGB 80303)." J Clin Oncol **28**(22): 3617-3622.

Kindler, H. L., K. Wroblewski, et al. (2012). "Gemcitabine plus sorafenib in patients with advanced pancreatic cancer: a phase II trial of the University of Chicago Phase II Consortium." Invest New Drugs **30**(1): 382-386.

Kogel, D., J. H. Prehn, et al. (2001). "The DAP kinase family of pro-apoptotic proteins: novel players in the apoptotic game." Bioessays **23**(4): 352-358.

Kong, L., Y. Zhang, et al. (2007). "CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine." Nucleic Acids Res **35**(Web Server issue): W345-349.

Kong, X., L. Li, et al. (2012). "Targeted destruction of the orchestration of the pancreatic stroma and tumor cells in pancreatic cancer cases: Molecular basis for therapeutic implications." Cytokine Growth Factor Rev **23**(6): 343-356.

Kreutzer, J. N., M. Ruzzene, et al. (2010). "Enhancing chemosensitivity to gemcitabine via RNA interference targeting the catalytic subunits of protein kinase CK2 in human pancreatic cancer cells." BMC Cancer **10**: 440.

Kutter, C., S. Watt, et al. (2012). "Rapid turnover of long noncoding RNAs and the evolution of gene expression." PLoS Genet **8**(7): e1002841.

Lages, E., H. Ipas, et al. (2012). "MicroRNAs: molecular features and role in cancer." Front Biosci **17**: 2508-2540.

Lander, E. S. (1999). "Array of hope." Nat Genet **21**(1 Suppl): 3-4.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lestrade, L. and M. J. Weber (2006). "snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs." Nucleic Acids Res **34**(Database issue): D158-162.

Li, L., T. Feng, et al. (2009). "Role of human noncoding RNAs in the control of tumorigenesis." Proc Natl Acad Sci U S A **106**(31): 12956-12961.

Liang, F., I. Holt, et al. (2000). "Gene index analysis of the human genome estimates approximately 120,000 genes." Nat Genet **25**(2): 239-240.

Liu, H., J. Shi, et al. (2012). "Reevaluation and identification of the best immunohistochemical panel (pVHL, Maspin, S100P, IMP-3) for ductal adenocarcinoma of the pancreas." Arch Pathol Lab Med **136**(6): 601-609.

Logsdon, C. D., D. M. Simeone, et al. (2003). "Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer." Cancer Res **63**(10): 2649-2657.

Lomnytska, M. I., S. Becker, et al. "Diagnostic protein marker patterns in squamous cervical cancer." Proteomics Clin Appl **4**(1): 17-31.

Lopez-Casas, P. P. and L. A. Lopez-Fernandez (2010). "Gene-expression profiling in pancreatic cancer." Expert Rev Mol Diagn **10**(5): 591-601.

Loukopoulos, P., T. Shibata, et al. (2007). "Genome-wide array-based comparative genomic hybridization analysis of pancreatic adenocarcinoma: identification of genetic indicators that predict patient outcome." Cancer Sci **98**(3): 392-400.

Louro, R., T. El-Jundi, et al. (2008). "Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci." Genomics **92**(1): 18-25.

Louro, R., H. I. Nakaya, et al. (2007). "Androgen responsive intronic non-coding RNAs." BMC Biol **5**: 4.

Lowe, A. W., M. Olsen, et al. (2007). "Gene expression patterns in pancreatic tumors, cells and tissues." PLoS One **2**(3): e323.

Lowenfels, A. B., P. Maisonneuve, et al. (1997). "Hereditary pancreatitis and the risk of pancreatic cancer. International Hereditary Pancreatitis Study Group." J Natl Cancer Inst **89**(6): 442-446.

Lowenfels, A. B., P. Maisonneuve, et al. (2001). "Cigarette smoking as a risk factor for pancreatic cancer in patients with hereditary pancreatitis." JAMA **286**(2): 169-170.

Lu, Z., L. Hu, et al. (2004). "Differential expression profiling of human pancreatic adenocarcinoma and healthy pancreatic tissue." Proteomics **4**(12): 3975-3988.

Lu, Z. J., K. Y. Yip, et al. (2010). "Prediction and characterization of noncoding RNAs in C. elegans by integrating conservation, secondary structure, and high-throughput sequencing and array data." Genome Res **21**(2): 276-285.

Luttges, J., B. Feyerabend, et al. (2002). "The mucin profile of noninvasive and invasive mucinous cystic neoplasms of the pancreas." Am J Surg Pathol **26**(4): 466-471.

Mahadevan, D. and D. D. Von Hoff (2007). "Tumor-stroma interactions in pancreatic ductal adenocarcinoma." Mol Cancer Ther **6**(4): 1186-1197.

Maitra, A., D. E. Hansel, et al. (2003). "Global expression analysis of well-differentiated pancreatic endocrine neoplasms using oligonucleotide microarrays." Clin Cancer Res **9**(16 Pt 1): 5988-5995.

Maitra, A. and R. H. Hruban (2008). "Pancreatic cancer." Annu Rev Pathol **3**: 157-188.

Manegold, C., J. Vansteenkiste, et al. (2012). "Randomized phase II study of three doses of the integrin inhibitor cilengitide versus docetaxel as second-line treatment for patients with advanced non-small-cell lung cancer." Invest New Drugs.

Mardis, E. R. (2012). "Applying next-generation sequencing to pancreatic cancer treatment." Nat Rev Gastroenterol Hepatol **9**(8): 477-486.

Martinez-Llordella, M., J. J. Lozano, et al. (2008). "Using transcriptional profiling to develop a diagnostic test of operational tolerance in liver transplant recipients." Journal of Clinical Investigation **118**(8): 2845-2857.

Masamune, A. and T. Shimosegawa (2009). "Signal transduction in pancreatic stellate cells." J Gastroenterol **44**(4): 249-260.

Matthaei, H., R. D. Schulick, et al. (2011). "Cystic precursors to invasive pancreatic cancer." Nat Rev Gastroenterol Hepatol **8**(3): 141-150.

Mattick, J. S. (2003). "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms." Bioessays **25**(10): 930-939.

Mattick, J. S. (2004). "RNA regulation: a new genetics?" Nat Rev Genet **5**(4): 316-323.

Mattick, J. S. (2009). "The genetic signatures of noncoding RNAs." PLoS Genet **5**(4): e1000459.

Matys, V., O. V. Kel-Margoulis, et al. (2006). "TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes." Nucleic Acids Res **34**(Database issue): D108-110.

McLaughlin-Drubin, M. E., C. P. Crum, et al. "Human papillomavirus E7 oncoprotein induces KDM6A and KDM6B histone demethylase expression and causes epigenetic reprogramming." Proc Natl Acad Sci U S A **108**(5): 2130-2135.

Mendoza-Villanueva, D., W. Deng, et al. "The Runx transcriptional co-activator, CBFbeta, is essential for invasion of breast cancer cells." Mol Cancer **9**: 171.

Mercer, T. R., M. E. Dinger, et al. (2008). "Specific expression of long noncoding RNAs in the mouse brain." Proc Natl Acad Sci U S A **105**(2): 716-721.

Mestdagh, P., E. Fredlund, et al. (2010). "An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours." Oncogene **29**(24): 3583-3592.

Mikkelsen, T. S., M. Ku, et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature **448**(7153): 553-560.

Mohammed, A., N. B. Janakiram, et al. (2012). "Early detection and prevention of pancreatic cancer: use of genetically engineered mouse models and advanced imaging technologies." Curr Med Chem **19**(22): 3701-3713.

Moore, M. J., D. Goldstein, et al. (2007). "Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group." J Clin Oncol **25**(15): 1960-1966.

Mortazavi, A., B. A. Williams, et al. (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods **5**(7): 621-628.

Moskaluk, C. A., R. H. Hruban, et al. (1997). "p16 and K-ras gene mutations in the intraductal precursors of human pancreatic adenocarcinoma." Cancer Res **57**(11): 2140-2143.

Nagai, M. A., R. Gerhard, et al. "Down-regulation of the candidate tumor suppressor gene PAR-4 is associated with poor prognosis in breast cancer." Int J Oncol **37**(1): 41-49.

Nagano, T. and P. Fraser (2011). "No-nonsense functions for long noncoding RNAs." Cell **145**(2): 178-181.

Nakagawa, H., S. Liyanarachchi, et al. (2004). "Role of cancer-associated stromal fibroblasts in metastatic colon cancer to the liver and their expression profiles." Oncogene **23**(44): 7366-7377.

Nakamura, T., Y. Furukawa, et al. (2004). "Genome-wide cDNA microarray analysis of gene expression profiles in pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for purity by laser microdissection." Oncogene **23**(13): 2385-2400.

Nakamura, T., T. Kuwai, et al. (2007). "Zonal heterogeneity for gene expression in human pancreatic carcinoma." Cancer Res **67**(16): 7597-7604.

Nakaya, H. I., P. P. Amaral, et al. (2007). "Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription." Genome Biol **8**(3): R43.

Neoptolemos, J. P., D. D. Stocken, et al. (2004). "A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer." N Engl J Med **350**(12): 1200-1210.

Nie, L., H. J. Wu, et al. (2012). "Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer." Am J Transl Res **4**(2): 127-150.

Numata, K., A. Kanai, et al. (2003). "Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection." Genome Res **13**(6B): 1301-1306.

Ohashi, Y., K. E. Creek, et al. (2004). "RNA degradation in human breast tissue after surgical removal: a time-course study." Exp Mol Pathol **77**(2): 98-103.

Ohuchida, K., K. Mizumoto, et al. (2007). "S100A6 is increased in a stepwise manner during pancreatic carcinogenesis: clinical value of expression analysis in 98 pancreatic juice samples." Cancer Epidemiol Biomarkers Prev **16**(4): 649-654.

Okazaki, Y., M. Furuno, et al. (2002). "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs." Nature **420**(6915): 563-573.

Olive, K. P., M. A. Jacobetz, et al. (2009). "Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer." Science **324**(5933): 1457-1461.

Ou, C. Y., J. H. Kim, et al. (2009). "Requirement of cell cycle and apoptosis regulator 1 for target gene activation by Wnt and beta-catenin and for anchorage-independent growth of human colon carcinoma cells." J Biol Chem **284**(31): 20629-20637.

Paez-Ribes, M., E. Allen, et al. (2009). "Antiangiogenic therapy elicits malignant progression of tumors to increased local invasion and distant metastasis." Cancer Cell **15**(3): 220-231.

Pan, S., R. Chen, et al. (2011). "Protein Alterations Associated with Pancreatic Cancer and Chronic Pancreatitis Found in Human Plasma using Global Quantitative Proteomics Profiling." J Proteome Res.

Pang, K. C., M. C. Frith, et al. (2006). "Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function." Trends Genet **22**(1): 1-5.

Panzitt, K., M. M. Tschernatsch, et al. (2007). "Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA." Gastroenterology **132**(1): 330-342.

Papaconstantinou, I. G., P. M. Lykoudis, et al. (2012). "A review on the role of microRNA in biology, diagnosis, and treatment of pancreatic adenocarcinoma." Pancreas **41**(5): 671-677.

Passerini, R., M. C. Cassatella, et al. (2012). "The pitfalls of CA19-9: routine testing and comparison of two automated immunoassays in a reference oncology center." Am J Clin Pathol **138**(2): 281-287.

Perez, D. S., T. R. Hoage, et al. (2008). "Long, abundantly expressed non-coding transcripts are altered in cancer." Hum Mol Genet **17**(5): 642-655.

Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Res **29**(9): e45.

Pilarsky, C., O. Ammerpohl, et al. (2008). "Activation of Wnt signalling in stroma from pancreatic cancer identified by gene expression profiling." J Cell Mol Med **12**(6B): 2823-2835.

Pleskow, D. K., H. J. Berger, et al. (1989). "Evaluation of a serologic marker, CA19-9, in the diagnosis of pancreatic cancer." Ann Intern Med **110**(9): 704-709.

Pogue-Geile, K. L., R. Chen, et al. (2006). "Palladin mutation causes familial pancreatic cancer and suggests a new cancer mechanism." PLoS Med **3**(12): e516.

Ponjavic, J., C. P. Ponting, et al. (2007). "Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs." Genome Res **17**(5): 556-565.

Prasad, N. B., A. V. Biankin, et al. (2005). "Gene expression profiles in pancreatic intraepithelial neoplasia reflect the effects of Hedgehog signaling on pancreatic ductal epithelial cells." Cancer Res **65**(5): 1619-1626.

Prensner, J. R. and A. M. Chinnaiyan (2011). "The emergence of lncRNAs in cancer biology." Cancer Discov **1**(5): 391-407.

Prensner, J. R., M. K. Iyer, et al. (2011). "Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression." Nat Biotechnol **29**(8): 742-749.

Quackenbush, J. (2002). "Microarray data normalization and transformation." Nat Genet **32 Suppl**: 496-501.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Ravasi, T., H. Suzuki, et al. (2006). "Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome." Genome Res **16**(1): 11-19.

Raycroft, L., H. Y. Wu, et al. (1990). "Transcriptional activation by wild-type but not transforming mutants of the p53 anti-oncogene." Science **249**(4972): 1049-1051.

Reddi, K. K. and J. F. Holland (1976). "Elevated serum ribonuclease in patients with pancreatic cancer." Proc Natl Acad Sci U S A **73**(7): 2308-2310.

Redston, M. S., C. Caldas, et al. (1994). "p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions." Cancer Res **54**(11): 3025-3033.

Reis, E. M., H. I. Nakaya, et al. (2004). "Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer." Oncogene **23**(39): 6684-6692.

Reis, E. M. and S. Verjovski-Almeida (2012). "Perspectives of Long Non-Coding RNAs in Cancer Diagnostics." Front Genet **3**: 32.

Rhim, A. D., E. T. Mirek, et al. (2012). "EMT and dissemination precede pancreatic tumor formation." Cell **148**(1-2): 349-361.

Rinn, J. L. and H. Y. Chang (2012). "Genome regulation by long noncoding RNAs." Annu Rev Biochem **81**: 145-166.

Rinn, J. L., G. Euskirchen, et al. (2003). "The transcriptional activity of human Chromosome 22." Genes Dev **17**(4): 529-540.

Rinn, J. L., M. Kertesz, et al. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell **129**(7): 1311-1323.

Ruckert, F., G. Dawelbait, et al. (2010). "Examination of apoptosis signaling in pancreatic cancer by computational signal transduction analysis." PLoS One **5**(8): e12243.

Saha, S., A. B. Sparks, et al. (2002). "Using the transcriptome to annotate the genome." Nat Biotechnol **20**(5): 508-512.

Salvia, R., C. Fernandez-del Castillo, et al. (2004). "Main-duct intraductal papillary mucinous neoplasms of the pancreas: clinical predictors of malignancy and long-term survival following resection." Ann Surg **239**(5): 678-685; discussion 685-677.

Sati, S., S. Ghosh, et al. (2012). "Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci." Nucleic Acids Res **40**(20): 10018-10031.

Sato, N., N. Fukushima, et al. (2004). "Gene expression profiling identifies genes associated with invasive intraductal papillary mucinous neoplasms of the pancreas." Am J Pathol **164**(3): 903-914.

Schubbert, S., K. Shannon, et al. (2007). "Hyperactive Ras in developmental disorders and cancer." Nat Rev Cancer **7**(4): 295-308.

Schutte, M., R. H. Hruban, et al. (1997). "Abrogation of the Rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas." Cancer Res **57**(15): 3126-3130.

Schwartz, S., W. J. Kent, et al. (2003). "Human-mouse alignments with BLASTZ." Genome Res **13**(1): 103-107.

Segara, D., A. V. Biankin, et al. (2005). "Expression of HOXB2, a retinoic acid signaling target in pancreatic cancer and pancreatic intraepithelial neoplasia." Clin Cancer Res **11**(9): 3587-3596.

Seitz, S., E. Korsching, et al. (2006). "Genetic background of different cancer cell lines influences the gene set involved in chromosome 8 mediated breast tumor suppression." Genes Chromosomes Cancer **45**(6): 612-627.

Selga, E., C. Oleaga, et al. (2009). "Networking of differentially expressed genes in human cancer cells resistant to methotrexate." Genome Med **1**(9): 83.

Sergeant, G., R. van Eijsden, et al. (2012). "Pancreatic cancer circulating tumour cells express a cell motility gene signature that predicts survival after surgery." BMC Cancer **12**(1): 527.

Shekouh, A. R., C. C. Thompson, et al. (2003). "Application of laser capture microdissection combined with two-dimensional electrophoresis for the discovery of differentially regulated proteins in pancreatic ductal adenocarcinoma." Proteomics **3**(10): 1988-2001.

Shen, J., M. D. Person, et al. (2004). "Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry." Cancer Res **64**(24): 9018-9026.

Sherman, B. T., W. Huang da, et al. (2007). "DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis." BMC Bioinformatics **8**: 426.

Siegel, R., E. Ward, et al. (2011). "Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths." CA Cancer J Clin **61**(4): 212-236.

Sierko, E., M. Z. Wojtukiewicz, et al. "Expression of tissue factor pathway inhibitor (TFPI) in human breast and colon cancer tissue." Thromb Haemost **103**(1): 198-204.

Silva, J. M., E. Ezhkova, et al. (2009). "Cyfip1 is a putative invasion suppressor in epithelial cancers." Cell **137**(6): 1047-1061.

Singh, M. and A. Maitra (2007). "Precursor lesions of pancreatic cancer: molecular pathology and clinical implications." Pancreatology **7**(1): 9-19.

Sohn, T. A., C. J. Yeo, et al. (2004). "Intraductal papillary mucinous neoplasms of the pancreas: an updated experience." Ann Surg **239**(6): 788-797; discussion 797-789.

Song, D., R. Chaerkady, et al. (2008). "Antitumor activity and molecular effects of the novel heat shock protein 90 inhibitor, IPI-504, in pancreatic cancer." Mol Cancer Ther **7**(10): 3275-3284.

St Laurent, G., D. Shtokalo, et al. (2012). "Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells." BMC Genomics **13**: 504.

Stapleton, G., A. Malliri, et al. (2002). "Downregulated AP-1 activity is associated with inhibition of Protein-Kinase-C-dependent CD44 and ezrin localisation and upregulation of PKC theta in A431 cells." J Cell Sci **115**(Pt 13): 2713-2724.

Staub, E., J. Groene, et al. (2009). "An expression module of WIPF1-coexpressed genes identifies patients with favorable prognosis in three tumor types." J Mol Med **87**(6): 633-644.

Steele, C. W., K. A. Oien, et al. (2011). "Clinical Potential of MicroRNAs in Pancreatic Ductal Adenocarcinoma." Pancreas **40**(8): 1165-1171.

Stevens, C., Y. Lin, et al. (2009). "Peptide combinatorial libraries identify TSC2 as a death-associated protein kinase (DAPK) death domain-binding protein and reveal a stimulatory role for DAPK in mTORC1 signaling." J Biol Chem **284**(1): 334-344.

Storey, J. D. and R. Tibshirani (2003). "Statistical significance for genomewide studies." Proc Natl Acad Sci U S A **100**(16): 9440-9445.

Stratford, J. K., D. J. Bentrem, et al. (2010). "A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma." PLoS Med **7**(7): e1000307.

Struhl, K. (2007). "Transcriptional noise and the fidelity of initiation by RNA polymerase II." Nat Struct Mol Biol **14**(2): 103-105.

Sugiyama, M., Y. Suzuki, et al. (2008). "Management of intraductal papillary mucinous neoplasm of the pancreas." J Gastroenterol **43**(3): 181-185.

Tabata, T., N. Tsukamoto, et al. (2009). "RNA interference targeting against S100A4 suppresses cell growth and motility and induces apoptosis in human pancreatic cancer cells." Biochem Biophys Res Commun.

Taft, R. J., E. A. Glazov, et al. (2009). "Tiny RNAs associated with transcription start sites in animals." Nat Genet **41**(5): 572-578.

Talamini, G., C. Bassi, et al. (1999). "Alcohol and smoking as risk factors in chronic pancreatitis and pancreatic cancer." Dig Dis Sci **44**(7): 1303-1311.

Tan, Z. J., X. G. Hu, et al. (2003). "Analysis of gene expression profile of pancreatic carcinoma using cDNA microarray." World J Gastroenterol **9**(4): 818-823.

Tano, K., R. Mizuno, et al. (2010). "MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes." FEBS Lett **584**(22): 4575-4580.

Team, R. D. C. (2011). "R: A language and environment for statistical computing." from http://www.R-project.org/.

Ting, D. T., D. Lipson, et al. (2011). "Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers." Science **331**(6017): 593-596.

Tinzl, M., M. Marberger, et al. (2004). "DD3PCA3 RNA analysis in urine--a new perspective for detecting prostate cancer." Eur Urol **46**(2): 182-186; discussion 187.

Tripathi, V., J. D. Ellis, et al. (2010). "The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation." Mol Cell **39**(6): 925-938.

Tsai, M. C., O. Manor, et al. (2010). "Long noncoding RNA as modular scaffold of histone modification complexes." Science **329**(5992): 689-693.

Tusher, V. G., R. Tibshirani, et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A **98**(9): 5116-5121.

Ueki, T., M. Toyota, et al. (2000). "Hypermethylation of multiple genes in pancreatic adenocarcinoma." Cancer Res **60**(7): 1835-1839.

Ulitsky, I., A. Shkumatava, et al. (2011). "Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution." Cell **147**(7): 1537-1550.

van Bakel, H., C. Nislow, et al. (2010). "Most "dark matter" transcripts are associated with known genes." PLoS Biol **8**(5): e1000371.

Van Gelder, R. N., M. E. von Zastrow, et al. (1990). "Amplified RNA synthesized from limited quantities of heterogeneous cDNA." Proc Natl Acad Sci U S A **87**(5): 1663-1667.

van Heek, N. T., A. K. Meeker, et al. (2002). "Telomere shortening is nearly universal in pancreatic intraepithelial neoplasia." Am J Pathol **161**(5): 1541-1547.

Vandesompele, J., K. De Preter, et al. (2002). "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes." Genome Biol **3**(7): RESEARCH0034.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.

Vimalachandran, D., W. Greenhalf, et al. (2005). "High nuclear S100A6 (Calcyclin) is significantly associated with poor survival in pancreatic cancer patients." Cancer Res **65**(8): 3218-3225.

Virtanen, C. and J. Woodgett (2008). "Clinical uses of microarrays in cancer research." Methods Mol Med **141**: 87-113.

Vlecken, D. H. and C. P. Bagowski (2009). "LIMK1 and LIMK2 are important for metastatic behavior and tumor cell-induced angiogenesis of pancreatic cancer cells." Zebrafish **6**(4): 433-439.

Vogelstein, B., D. Lane, et al. (2000). "Surfing the p53 network." Nature **408**(6810): 307-310.

Wang, E., L. D. Miller, et al. (2000). "High-fidelity mRNA amplification for gene profiling." Nat Biotechnol **18**(4): 457-459.

Wang, E. T., R. Sandberg, et al. (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-476.

Wang, L., S. Tsutsumi, et al. (2012). "Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency." Genome Res **22**(2): 208-219.

Wang, Y., R. Gangeswaran, et al. (2009). "CEACAM6 attenuates adenovirus infection by antagonizing viral trafficking in cancer cells." J Clin Invest **119**(6): 1604-1615.

Wang, Z., C. Zang, et al. (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." Nat Genet **40**(7): 897-903.

Washietl, S., I. L. Hofacker, et al. (2005). "Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome." Nat Biotechnol **23**(11): 1383-1390.

Weichert, W., M. Boehm, et al. (2007). "High expression of RelA/p65 is associated with activation of nuclear factor-kappaB-dependent signaling in pancreatic cancer and marks a patient population with poor prognosis." Br J Cancer **97**(4): 523-530.

Whitcomb, D. C. (2004). "Inflammation and Cancer V. Chronic pancreatitis and pancreatic cancer." Am J Physiol Gastrointest Liver Physiol **287**(2): G315-319.

Willingham, A. T. and T. R. Gingeras (2006). "TUF love for "junk" DNA." Cell **125**(7): 1215-1220.

Wilusz, J. E., S. M. Freier, et al. (2008). "3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA." Cell **135**(5): 919-932.

Wilusz, J. E., H. Sunwoo, et al. (2009). "Long noncoding RNAs: functional surprises from the RNA world." Genes Dev **23**(13): 1494-1504.

Winter, J. M., C. J. Yeo, et al. (2012). "Diagnostic, prognostic, and predictive biomarkers in pancreatic cancer." J Surg Oncol.

Wright, M. W. and E. A. Bruford (2011). "Naming 'junk': human non-protein coding RNA (ncRNA) gene nomenclature." Hum Genomics **5**(2): 90-98.

Wu, C. H., R. Apweiler, et al. (2006). "The Universal Protein Resource (UniProt): an expanding universe of protein information." Nucleic Acids Res **34**(Database issue): D187-191.

Wu, Z. Z., H. P. Lu, et al. "Identification and functional analysis of genes which confer resistance to cisplatin in tumor cells." Biochem Pharmacol **80**(2): 262-276.

Yachida, S. and C. A. Iacobuzio-Donahue (2009). "The pathology and genetics of metastatic pancreatic cancer." Arch Pathol Lab Med **133**(3): 413-422.

Yachida, S., S. Jones, et al. (2010). "Distant metastasis occurs late during the genetic evolution of pancreatic cancer." Nature **467**(7319): 1114-1117.

Yamaguchi, K., S. Kanemitsu, et al. (2011). "Pancreatic ductal adenocarcinoma derived from IPMN and pancreatic ductal adenocarcinoma concomitant with IPMN." Pancreas **40**(4): 571-580.

Yan, M. D., C. C. Hong, et al. (2005). "Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas." Hum Mol Genet **14**(11): 1465-1474.

Yap, K. L., S. Li, et al. (2010). "Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a." Mol Cell **38**(5): 662-674.

Yin, Z., J. Sun, et al. (2012). "Surgical treatment strategies in chronic pancreatitis: a meta-analysis." Arch Surg **147**(10): 961-968.

Ying, L., Q. Chen, et al. (2012). "Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition." Mol Biosyst **8**(9): 2289-2294.

Yokoyama, Y., Y. Nimura, et al. (2009). "Advances in the treatment of pancreatic cancer: limitations of surgery and evaluation of new therapeutic strategies." Surg Today **39**(6): 466-475.

Yu, K. H., A. K. Rustgi, et al. (2005). "Characterization of proteins in human pancreatic cancer serum using differential gel electrophoresis and tandem mass spectrometry." J Proteome Res **4**(5): 1742-1751.

Yu, W., D. Gius, et al. (2008). "Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA." Nature **451**(7175): 202-206.

Zhang, H., L. Lv, et al. (2009). "Profiling the potential biomarkers for cell differentiation of pancreatic cancer using iTRAQ and 2-D LC-MS/MS." Proteomics Clin Appl **3**(7): 862-871.

Zhang, L., J. J. Farrell, et al. (2009). "Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer." Gastroenterology **138**(3): 949-957 e941-947.

Zhang, X., K. Rice, et al. (2010). "Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions." Endocrinology **151**(3): 939-947.

Zhao, Y., S. Shen, et al. (2006). "Mitogen-activated protein kinases and chemoresistance in pancreatic cancer cells." J Surg Res **136**(2): 325-335.

Zhu, J., J. Z. Sanborn, et al. (2007). "Comparative genomics search for losses of long-established genes on the human lineage." PLoS Comput Biol **3**(12): e247.

APÊNDICE

Apêndice 1: Tabela de *primers* utilizados no projeto

Primers	Sequencia *Foward*	Sequencia *Reverse*
ARRB1	TTCAGCCTGCCCCAGAACT	GCTCTCATCCCACAGCAGAGA
ATF2	ATCAAAGAAGGTGAAATGAAAAAGC	CAGTGTAGCTGTATTGGTGAGATTATTG
MAP2K1	GAGACATACCACTAAGTATCCCACACA	GGGCTTCACCAGTCTGTCTACTG
MAP3K1	TGCACAATGCAAACTGAAATCA	GCAATGGAGGGAATGTCCTACT
MAP3K14	CGGTCAATCTCGCTGAAATACA	ACACAGGTAGACGCTGCTTGAA
MAPK1	TCTGACTGATGCGGCTATGG	CAATTTAGAGAGCAAAGAGGTTGAGA
MAPK3	CTTGGAGGGCAGAGACTGTAGGT	CAACCACATTCTGGGCATCCT
PPP3CB	AGAAAGTATAAACTGACTCCATGCCA	CTGGTTTTGAAATGTAGAAGTGTGC
RAPGF2	CCCAAACCCAACCTTTTTTGT	TCGGGAAATGAGTGTTTAAAGTCA
DAPK1	TCTGACCCGTGCACCTACCT	GTCTCAAACAATCACACACCAACTC
PTEN		
HMBS	GGCAATGCGGCTGCAA	GGGTACCCACGCGAATCAC
DAPK-1 exonic	GCATCCTGAAGGAGATCCAG	GATTTGAAGGGAGTGCAGGT
MAPK3K14_exonic	CGGGTCAAAGTGGGAGACA	CGAGTCTGGCACCTCCATGT
PPP3CB_exonic	CCTGACTCCCACAGGGATGTT	GCAGATTCTATGTGGTGGAGAGAA
DAPK1_RACE	GATGATGTGGAAAATCGGAGTGTCA	CGTGCACCTACCTGCCACTGA
DAPK1_NESTED	TCTTCCACCTTCTTCTAACTGCACCTT	TGCATCCTTTCTCGGCCTTG
PPP3CB_RACE	CCCCTCAGCCAGACACAAGTGG	GGTGACGTTTTTCTACCACTCAACCTG
PPP3CB_NESTED	TGTTTTCACGCACACTGTGAATGCTC	GCTGTCCCAGCTGGTTTTGAAATGT
MAP3K14_RACE	GTGGGGGCTGCCCTGTATGC	GGGAGCCTGAGCCTGTCTTTCTTCA
MAP3K14_NESTED	CCACCCACCGCTCCTTTCTGG	GCCACTCCCACCTCTGAAAACCA
PPP3CB_1200	TGTGTTCTACATGATGAATATTTCC	CAACTTATTTTAGGCCAAATTCT
DAPK1_1050	GGACTTTGACCTGGAAGAAG	TTGTTTTGGATTGGGACAG
DAPK1_long_isoform	GGCACCCTCATGTCCAAACT	CATTGCCATCCAGGTTGATTT
DAPK1_promotor_senso_longo	ATAAGCTAGCGGCCAGGAACAGAGTAGGTG	ATAAAAGCTTACTCTTCCCCAATGGAAACC
DAPK1_promotor_antisenso_longo	ATAAAAGCTTGGCCAGGAACAGAGTAGGTG	ATAAGCTAGCACTCTTCCCCAATGGAAACC
DAPK1_promotor_senso_curto	GGTAGCTAGCAGTGCACTCATCCCTCGGTA	GCTAAAGCTTTCCATGGATTCAACCAACCT
DAPK1_promotor_antisenso_curto	GCTAAAGCTTAGTGCACTCATCCCTCGGTA	GTTAGCTAGCTCCATGGATTCAACCAACCT
pre1-DAPK1	AGAGATCAGTTCTGTGTGTTCCCTA	GGCTCTACTGTTAGGCTCTGAGAA
pre2-DAPK1	TGACTGGCATAGTGAGGAGTAGAG	GAAGAAAAGAAAATCCCAAACTGA
pre4-DAPK1	CCTGTGGTCATTTACGTCTCATAA	GAGGAAGCAAACTAACCGATTCTA

Apêndice 2: Tabela de genes diferencialmente expressos comparando tumores primário (T) com pancreatite crônica (CP) e tecidos não tumorais (NT).

Probe ID	Gene Bank Acession	Probe Region	Gene Name	Gene ID	Gene Symbol	Ref Seq Gene	Chromossome	Log_2 (Fold Change T/PC+NT)
3634	BF853527	Exonic	metallothionein 1H	4502	MT2A	NM_005953	chr16: 56642510 - 56643346	-1.25
4380	BQ358748	Exonic	thioesterase superfamily member 5	284486	THEM5	NM_182578	chr1: 151823547 - 151826116	-1.23
3964	AW748714	Exonic	zinc finger and SCAN domain containing 18	65982	ZSCAN18	NM_023926	chr19: 58599594 - 58600025	-1.14
1414	BE170357	Exonic	cell cycle progression 1	9236	CCPG1	NM_020739	chr15: 55652546 - 55652915	-1.05
4418	AW880616	Exonic	neuroligin 4, X-linked	57502	NLGN4X	NM_181332	chrX: 5808433 - 5808972	-1.02
1442	BF890756	Exonic	cullin 4B	8450	CUL4B	NM_003588	chrX: 119677966 - 119681021	-0.96
2945	BE813923	Exonic	AT rich interactive domain 4A (RBP1-like)	5926	ARID4A	NM_023001	chr14: 58838692 - 58839233	-0.94
3163	BE080775	Exonic	similar to transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila); transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	7088	TLE1	NM_005077	chr9: 84248419 - 84248751	-0.91
3994	BF928925	Exonic	leucine rich repeat containing 57	255252	LRRC57	NM_153260	chr15: 42836092 - 42839500	-0.91
4116	BF956998	Exonic	Notch homolog 4 (Drosophila)	4855	NOTCH4	NM_004557	chr6: 32180251 - 32182014	-0.89
1289	BF883365	Exonic	MYC induced nuclear antigen	84864	MINA	NM_153182	chr3: 97664008 - 97668764	-0.88
2870	BF920836	Exonic	transcription factor binding to IGHM enhancer 3	7030	TFE3	NM_006521	chrX: 48886720 - 48887130	-0.87
4109	AW902244	Exonic	CaM kinase-like vesicle-associated	79012	CAMKV	NM_024046	chr3: 49895715 - 49896156	-0.83
2564	BG003639	Exonic	neural cell adhesion molecule 1	4684	NCAM1	NM_000615	chr11: 113076801 - 113078666	-0.83
4157	BE161600	Exonic	ST3 beta-galactoside alpha-2,3-sialyltransferase 5	8869	ST3GAL5	NM_003896	chr2: 86073499 - 86075194	-0.80
1377	BF907266	Exonic	THO complex 2	57187	THOC2	NM_001081550	chrX: 122737000 - 122737309	-0.74
4002	BF816617	Exonic	retinoid X receptor, alpha	6256	RXRA	NM_002957	chr9: 137331492 - 137331855	-0.74
3098	BF848512	Exonic	WAS/WASL interacting protein family, member 2	147179	WIPF2	NM_133264	chr17: 38438372 - 38438743	-0.74
4045	AW383781	Exonic	tripartite motif-containing 26	7726	TRIM26	NM_003449	chr6: 30152380 - 30152901	-0.71
3649	BF889142	Exonic	CREB binding protein	1387	CREBBP	NM_001079846	chr16: 3776899 - 3777317	-0.70
2083	BE708499	Exonic	deleted in liver cancer 1	10395	DLC1	NM_182643	chr8: 12943203 - 12943976	-0.69

3674	BE694434	Exonic	A kinase (PRKA) anchor protein 11	11215	AKAP11	NM_016248	chr13: 42877859 - 42891701	-0.67
4354	BF997721	Exonic	tripartite motif-containing 25	7706	TRIM25	NM_005082	chr17: 54966236 - 54966713	-0.66
1615	BE926631	Exonic	ankyrin repeat domain 40	91369	ANKRD40	NM_052855	chr17: 48772971 - 48773355	-0.65
4237	BE089334	Exonic	eukaryotic translation initiation factor 4E binding protein 2	1979	EIF4EBP2	NM_004096	chr10: 72185273 - 72185618	-0.65
4127	BE825688	Exonic	forkhead box N2	3344	FOXN2	NM_002158	chr2: 48573374 - 48573739	-0.65
3719	BQ318852	Exonic	eukaryotic translation initiation factor 1A, X-linked	9086	EIF1AY	NM_004681	chrY: 22744500 - 22754371	-0.62
993	AW748486	Exonic	StAR-related lipid transfer (START) domain containing 3	10948	STARD3	NM_001165937	chr17: 37819074 - 37819526	-0.61
4079	BF083029	Exonic	nuclear mitotic apparatus protein 1	4926	NUMA1	NM_006185	chr11: 71714302 - 71714553	-0.60
3348	BF956997	Exonic	insulin receptor	3643	INSR	NM_001079817	chr19: 7115432 - 7116001	-0.59
4479	BF851788	Exonic	metastasis associated 1	9112	MTA1	NM_004689	chr14: 105935806 - 105936531	-0.59
4358	BE163446	Exonic	solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 23	79085	SLC25A23	NM_024103	chr19: 6440350 - 6440918	-0.59
3341	AW902178	Exonic	coronin, actin binding protein, 2B	10391	CORO2B	NM_001190456	chr15: 69019629 - 69019996	-0.58
983	AW352153	Exonic	PRKC, apoptosis, WT1, regulator	5074	PAWR	NM_002583	chr12: 79986375 - 79990605	-0.57
4016	BF156016	Exonic	formin binding protein 4	23360	FNBP4	NM_015308	chr11: 47738099 - 47738357	-0.56
1049	BE168033	Exonic	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	4790	NFKB1	NM_001165412	chr4: 103500103 - 103514719	-0.55
2160	BE832971	Exonic	phenylalanyl-tRNA synthetase, beta subunit	10056	FARSB	NM_005687	chr2: 223464681 - 223488446	-0.54
3892	AW880836	Exonic	neuropilin 2	8828	NRP2	NM_003872	chr2: 206661782 - 206662220	-0.54
3152	BF351473	Exonic	chromosome 1 open reading frame 115	79762	C1orf115	NM_024709	chr1: 220870418 - 220870743	-0.54
2159	BG009760	Exonic	T-cell leukemia translocation altered gene	6988	TCTA	NM_022171	chr3: 49453065 - 49453401	-0.54
2866	BF854041	Exonic	transforming, acidic coiled-coil containing protein 2	10579	TACC2	NM_006997	chr10: 124008155 - 124013595	-0.53
784	BE828277	Exonic	THO complex 7 homolog (Drosophila)	80145	THOC7	NM_025075	chr3: 63819755 - 63822036	-0.52
3334	BG005931	Exonic	breast cancer anti-estrogen resistance 3	8412	BCAR3	NM_003567	chr1: 94054875 - 94140324	-0.51
3919	BF082553	Exonic	RWD domain containing 4A	201965	RWDD4	NM_152682	chr4: 184560966 - 184561269	-0.49

4123	AW884452	Exonic	ribosomal protein L36a pseudogene 51; ribosomal protein L36a pseudogene 37; ribosomal protein L36a pseudogene 49; heterogeneous nuclear ribonucleoprotein H2 (H'); ribosomal protein L36a	6166	RPL36AL	NM_001001	chr14: 50085411 - 50085859	-0.48
3908	BF996163	Exonic	major facilitator superfamily domain containing 2	84879	MFSD2A	NM_001136493	chr1: 40432298 - 40434034	-0.47
3117	BE156258	Exonic	eukaryotic translation elongation factor 1 alpha-like 7; eukaryotic translation elongation factor 1 alpha-like 3; similar to eukaryotic translation elongation factor 1 alpha 1; eukaryotic translation elongation factor 1 alpha 1	645715	EEF1A1	NM_001402	chr6: 74225045 - 74225957	-0.47
674	BF890754	Exonic	ubiquinol-cytochrome c reductase hinge protein-like; ubiquinol-cytochrome c reductase hinge protein	7388	UQCRH	NM_006004	chr1: 46775825 - 46782443	-0.45
3720	AW889200	Exonic	similar to Down-regulated in metastasis protein (Key-1A6 protein) (Novel nucleolar protein 73) (NNP73); UTP20, small subunit (SSU) processome component, homolog (yeast)	27340	UTP20	NM_014503	chr12: 101779369 - 101779816	-0.45
2580	BF956989	Exonic	A kinase (PRKA) anchor protein 2; paralemmin 2/ PALM2-AKAP2 readthrough transcript		AKAP2/ PALM2-AKAP2	NM_001004065, NM_001136562, NM_001198656/NM_007203, NM_147150	chr9: 112900635 - 112900980	-0.44
3158	BG000783	Exonic	protease, serine, 12 (neurotrypsin, motopsin)	8492	PRSS12	NM_003619	chr4: 119201681 - 119202232	-0.44
4121	BE168342	Exonic	deformed epidermal autoregulatory factor 1 (Drosophila)	10522	DEAF1	NM_021008	chr11: 679719 - 687941	-0.44
1673	BF336884	Exonic	THO complex 3	84321	THOC3	NM_032361	chr5: 177309413 - 177310913	-0.43
2935	BG000772	Exonic	CD14 molecule	929	CD14	NM_000591	chr5: 140011485 - 140011954	-0.41
4507	BF756552	Exonic	axin 2	8313	AXIN2	NM_004655	chr17: 63525549 - 63525871	-0.39
3479	BE061111	Exonic	TAF8 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 43kDa	129685	TAF8	NM_138572	chr6: 42046925 - 42047215	-0.37
1608	BG001145	Exonic	nucleoporin 85kDa	79902	NUP85	NM_024844	chr17: 73224852 - 73225287	-0.37
873	CK327135	Exonic	synaptojanin 2 binding protein	55333	SYNJ2BP	NM_018373	chr14: 70837736 - 70838099	-0.32

2597	BF372924	Exonic	centrosomal protein 70kDa	80321	CEP70	NM_024491	chr3: 138224250 - 138248240	-0.30
1429	BF090391	Exonic	small proline-rich protein 1B	6699	SPRR1B	NM_003125	chr1: 153004798 - 153005343	-0.19
609	CK327153	Exonic	actin, gamma 1	71	ACTG1	NM_001614	chr17: 79477178 - 79477612	0.38
222	AW369265	Exonic	excision repair cross-complementing rodent repair deficiency, complementation group 5	2073	ERCC5	NM_000123	chr13: 103524547 - 103527941	0.45
1693	AW605304	Exonic	coiled-coil domain containing 52	152185	SPICE1	NM_144718	chr3: 113165456 - 113172534	0.60
2591	BE762975	Exonic	centrin, EF-hand protein, 2	1069	CETN2	NM_004344	chrX: 151996148 - 151997119	0.63
3583	AW805129	Exonic	E74-like factor 3 (ets domain transcription factor, epithelial-specific)	1999	ELF3	NM_001114309	chr1: 201985924 - 201986245	0.64
1408	BQ352893	Exonic	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	506	ATP5B	NM_001686	chr12: 57036569 - 57038715	0.66
1404	BE709116	Exonic	cell division cycle 25 homolog B (S. pombe)	994	CDC25B	NM_021873	chr20: 3786443 - 3786685	0.68
2518	BF988663	Exonic	syndecan 1	6382	SDC1	NM_002997	chr2: 20401873 - 20402387	0.71
251	BE826656	Exonic	thyroid hormone receptor interactor 6	7205	TRIP6	NM_003302	chr7: 100468210 - 100470982	0.72
928	BE841633	Exonic	membrane bound O-acyltransferase domain containing 2	129642	MBOAT2	NM_138799	chr2: 8998412 - 8998698	0.73
1375	BE166709	Exonic	integrin, beta 5	3693	ITGB5	NM_002213	chr3: 124483236 - 124492611	0.74
578	AW880912	Exonic	nischarin	11188	NISCH	NM_007184	chr3: 52526352 - 52526892	0.76
617	BF735925	Exonic	OCIA domain containing 1	54940	OCIAD1	NM_001079839	chr4: 48859228 - 48863139	0.78
601	BQ349514	Exonic	v-ral simian leukemia viral oncogene homolog A (ras related)	5898	RALA	NM_005402	chr7: 39726323 - 39736461	0.83
2193	BE833209	Exonic	N-myc (and STAT) interactor	9111	NMI	NM_004688	chr2: 152127313 - 152132079	0.84
1374	BE840232	Exonic	tumor suppressor candidate 3	7991	TUSC3	NM_006765	chr8: 15531254 - 15605963	0.91
1849	CK326978	Exonic	diacylglycerol kinase, theta 110kDa	1609	DGKQ	NM_001347	chr4: 954420 - 955282	0.93
3153	AW804421	Exonic	centromere protein T	80152	CENPT	NM_025082	chr16: 67863961 - 67864302	0.95
2975	BF883034	Exonic	actin related protein 2/3 complex, subunit 1B, 41kDa; similar to Actin-related protein 2/3 complex subunit 1B (ARP2/3 complex 41 kDa subunit) (p41-ARC)	10095	ARPC1B	NM_005720	chr7: 98985809 - 98990358	0.96
1600	AW393928	Exonic	transmembrane protein 87B	84910	TMEM87B	NM_032824	chr2: 112876058 - 112876711	0.96

4404	BF989969	Exonic	pituitary tumor-transforming 1 interacting protein	754	PTTG1IP	NM_004339	chr21: 46269711 - 46270281	0.96
287	BE833289	Exonic	vinculin	7414	VCL	NM_014000	chr10: 75879048 - 75879381	0.97
1043	AW605349	Exonic	transforming growth factor, beta-induced, 68kDa	7045	TGFBI	NM_000358	chr5: 135394777 - 135399121	1.01
642	BF883122	Exonic	CD63 molecule	967	CD63	NM_001040034	chr12: 56119916 - 56121037	1.04
2348	BF375449	Exonic	trans-golgi network protein 2	10618	TGOLN2	NM_006464	chr2: 85545357 - 85545693	1.05
3339	BQ310838	Exonic	heat shock protein 90kDa beta (Grp94), member 1	7184	HSP90B1	NM_003299	chr12: 104336329 - 104340474	1.05
1367	AW391467	Exonic	interferon induced transmembrane protein 1 (9-27)	8519	IFITM3	NM_021034	chr11: 319884 - 320757	1.08
1302	BE163603	Exonic	transmembrane protein 141	85014	TMEM141	NM_032928	chr9: 139686710 - 139687665	1.09
4431	BE166669	Exonic	beta-2-microglobulin	567	B2M	NM_004048	chr15: 45003766 - 45010267	1.15
2123	BQ349504	Exonic	lactate dehydrogenase A	3939	LDHA	NM_005566, NM_001135269, NM_001165414, NR_028500	chr11: 18427006 - 18429090	1.15
2143	AW819761	Exonic	cadherin 1, type 1, E-cadherin (epithelial)	999	CDH1	NM_004360	chr16: 68868514 - 68869081	1.17
229	AW939042	Exonic	S100 calcium binding protein A6	6277	S100A6	NM_014624	chr1: 153507124 - 153507833	1.20
558	BF825996	Exonic	PERP, TP53 apoptosis effector	64065	PERP	NM_022121	chr6: 138412333 - 138412921	1.21
1666	AW608129	Exonic	collagen triple helix repeat containing 1	115908	CTHRC1	NM_138455	chr8: 104388147 - 104394802	1.23
613	BF951773	Exonic	RAN, member RAS oncogene family	5901	RAN	NM_006325	chr12: 131357147 - 131360274	1.24
2157	BF880950	Exonic	hepsin	3249	HPN	NM_182983	chr19: 35556192 - 35557298	1.25
3270	BF985067	Exonic	collagen, type I, alpha 1	1277	COL1A1	NM_000088	chr17: 48261490 - 48261694	1.34
624	BE712464	Exonic	syntaxin 6	10228	STX6	NM_005819	chr1: 180959119 - 180974600	1.37
2226	AW852323	Exonic	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2	11014	KDELR2	NM_006854	chr7: 6500778 - 6501263	1.44
992	AW606067	Exonic	cyclin-dependent kinase 2 associated protein 1	8099	CDK2AP1	NM_004642	chr12: 123745864 - 123746351	1.48
3249	BF820605	Exonic	actin related protein 2/3 complex, subunit 2, 34kDa	10109	ARPC2	NM_152862	chr2: 219093458 - 219110266	1.74
2468	BF930663	Exonic	S100 calcium binding protein P	6286	S100P	NM_005980	chr4: 6698701 - 6699262	2.57
1742	AW370679	Exonic	TIMP metallopeptidase inhibitor 1	7076	TIMP1	NM_003254	chrX: 47444333 - 47446082	2.62

1670	AW902314	Exonic/ ncRNA	family with sequence similarity 36, member A/ non-protein coding RNA 201 (NCRNA00201)	116228/ 284702	FAM36A/ NCRNA00201	NM_198076/ NR_266778	chr1: 245007377 - 245007734	-0.69
4283	AW901940	Intergenic	-	-	-	-	chr15: 93442037 - 93442479	-0.93
3977	BF754591	Intergenic	-	-	-	-	chr1: 46653707 - 46654108	-0.91
3897	BE155946	Intergenic	-	-	-	-	chr20: 1025153 - 1025834	-0.73
3865	BE008094	Intergenic	-	-	-	-	chr2: 181710096 - 181710451	-0.68
4262	BE702744	Intergenic	-	-	-	-	chr7: 127944519 - 127944846	-0.67
1018	BF956245	Intergenic	-	-	-	-	chr7: 99345995 - 99346500	-0.60
4268	BF918966	Intergenic	-	-	-	-	chr5: 92193518 - 92194056	-0.56
2848	BF736784	Intergenic	-	-	-	-	chr1: 35644977 - 35645390	-0.55
4280	BF944704	Intergenic	-	-	-	-	chr1: 11353583 - 11355287	-0.42
4379	BF368581	Intronic	Sorbin and SH3 domain containing 1	10580	SORBS1	NM_015385	chr10: 97315461 - 97315809	-1.53
4360	BE168353	Intronic	PAS domain containing serine/threonine kinase	23178	PASK	NM_015148	chr2: 242086059 - 242086526	-1.05
4352	BE003587	Intronic	Protein phosphatase 4, regulatory subunit 1	9989	PPP4R1	NM_001042388	chr18: 9561281 - 9561760	-1.00
1168	AW852151	Intronic	DIP2 disco-interacting protein 2 homolog B (Drosophila)	57609	DIP2B	NM_173602	chr12: 51007113 - 51007544	-0.97
3842	BF334142	Intronic	Enolase superfamily member 1	55556	ENOSF1	NM_001126123	chr18: 709089 - 709430	-0.85
914	BF893298	Intronic	Mitogen-activated protein kinase kinase 5	5607	MAP2K5	NM_002757	chr15: 67889806 - 67890177	-0.82
4319	CK327106	Intronic	Tyrosine kinase, non-receptor, 2	10188	TNK2	NM_005781	chr3: 195591793 - 195592189	-0.80
3912	BG004239	Intronic	neuroblastoma breakpoint family, member 10	100132406	NBPF10	NM_001039703	chr1: 145302128 - 145302563	-0.76
3229	AW937483	Intronic	Activating signal cointegrator 1 complex subunit 3	10973	ASCC3	NM_006828	chr6: 101180631 - 101181205	-0.76
4339	BF898694	Intronic	Megakaryoblastic leukemia (translocation) 1	57591	MKL1	NM_020831	chr22: 40849141 - 40849644	-0.70
858	BG010309	Intronic	DAZ associated protein 1	26528	DAZAP1	NM_170711	chr19: 1419237 - 1419769	-0.69
4258	BF364131	Intronic	Chromosome 16 open reading frame 46	123775	C16orf46	NM_152337	chr16: 81104689 - 81105025	-0.67
3513	AW901803	Intronic	Gliomedin	342035	GLDN	NM_181789	chr15: 51642970 - 51643317	-0.66
4064	BQ316935	Intronic	Phosphonoformate immuno-associated protein 5	10443	N4BP2L2	NM_033111	chr13: 33109399 - 33109744	-0.64
76	BE160056	Intronic	Na+/K+ transporting ATPase interacting 2	154215	NKAIN2	NM_001040214	chr6: 124980949 - 124981492	-0.64
3886	AW820168	Intronic	Syntaxin 12	23673	STX12	NM_177424	chr1: 28141725 - 28142251	-0.63
2838	BE163917	Intronic	Guanine nucleotide binding protein (G protein), beta polypeptide 1	2782	GNB1	NM_002074	chr1: 1803765 - 1804063	-0.63

3903	BE166370	Intronic	Nuclear receptor subfamily 2, group C, member 2	7182	NR2C2	NM_003298	chr3: 15084054 - 15084337	-0.62
2435	AW846670	Intronic	Chromosome 11 open reading frame 49	79096	C11orf49	NM_001003676	chr11: 47169433 - 47169741	-0.61
4253	BE089575	Intronic	Chromosome 20 open reading frame 7	79133	C20orf7	NM_001039375	chr20: 13789721 - 13790060	-0.58
912	AW865910	Intronic	Nucleolar protein NOP5/NOP58	51602	NOP58	NM_015934	chr2: 203164022 - 203164376	-0.57
3979	CK327186	Intronic	Leptin receptor	3953	LEPR	NM_002303	chr1: 66071979 - 66072278	-0.55
790	AW860958	Intronic	Dystonin	667	DST	NM_015548	chr6: 56329341 - 56329893	-0.53
3585	BF875569	Intronic	Activating transcription factor 3	467	ATF3	NM_001030287	chr1: 212785642 - 212786212	-0.50
2384	BE168993	Intronic	Macrophage erythroblast attacher	10296	MAEA	NM_001017405	chr4: 1318234 - 1318651	-0.47
846	BE160464	Intronic	Gephyrin	10243	GPHN	NM_001024218	chr14: 66990664 - 66991047	-0.45
3849	BE007885	Intronic	Rho GTPase activating protein 12	94134	ARHGAP12	NM_018287	chr10: 32168422 - 32168792	-0.45
2411	BF894929	Intronic	Ubiquitin specific peptidase 15	9958	USP15	NM_006313	chr12: 62725137 - 62725661	-0.41
1194	BF944729	Intronic	Arrestin, beta 1	408	ARRB1	NM_020251	chr11: 75045505 - 75045972	-0.39
2784	BF857841	Intronic	Inositol 1,4,5-triphosphate receptor, type 1	3708	ITPR1	NM_001099952	chr3: 4786881 - 4787338	-0.33
4357	CK327190	Intronic	Zinc finger protein 644	84146	ZNF644	NM_016620	chr1: 91432941 - 91433257	-0.33
446	BF919473	Intronic	Catenin (cadherin-associated protein), delta 1	1500	CTNND1	NM_001085458	chr11: 57541563 - 57541976	-0.26
1695	AW937741	Intronic	SFRS protein kinase 1	6732	SRPK1	NM_003137	chr6: 35819568 - 35820194	1.00
3483	BE062809	Intronic/ Intronic	TMEM189-UBE2V1 readthrough/ ubiquitin-conjugating enzyme E2 variant 1	387522/ 7335	TMEM189-UBE2V1/ UBE2V1	NM_199203/ NM_001032288, NM_021988, NM_022442, NM_199144	chr20: 48718697 - 48719099	-0.57

Apêndice 3: Análise de enriquecimento (DAVID) de genes diferencialmente expressos na assinatura de amostras de tumores primários comparados à pancreatite crônica e amostras não tumorais.

Categoria	Termos	p-valor*	Genes
Genes codificadores de proteína			
GOTERM_BP_FAT GO:0019048	Virus-Host Interaction	0,005	THOC7, THOC2, INSR, THOC3
GOTERM_BP_FAT GO:0050658	RNA transport	0,007	THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0006405	RNA Export from Nucleus	0,005	THOC7, THOC2, RAN, THOC3
GOTERM_BP_FAT GO:0050657	Nucleic Acid Transport	0,007	THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0051236	Establishment of RNA Localization	0,007	THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0051168	Nuclear Export	0,017	THOC7, THOC2, RAN, THOC3
GOTERM_BP_FAT GO:0051028	mRNA Transport	0,035	THOC7, THOC2, NUP85, THOC3
GOTERM_BP_FAT GO:0006406	mRNA Export from Nucleus	0,039	THOC7, THOC2, THOC3
GOTERM_BP_FAT GO:0006403	RNA Localization	0,010	THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0015931	Nucleobase, containing compound Transport	0,010	THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0046784	Intronless Viral mRNA Export from Host Nucleus	0,021	THOC7, THOC2, THOC3
GOTERM_CC_FAT GO:0005925	Focal Adhesion	0,024	TRIP6, TRIM25, VCL, SDC1, ARPC2, DLC1, CDH1
GOTERM_CC_FAT GO:0030055	Cell-Substrate Junction	0,024	TRIP6, TRIM25, VCL, SDC1, ARPC2, DLC1, CDH1
GOTERM_CC_FAT GO:0005924	Cell-Substrate Adherens Junction	0,024	TRIP6, TRIM25, VCL, SDC1, ARPC2, DLC1, CDH1
GOTERM_CC_FAT GO:0070161	Anchoring Junction	0,044	TRIP6, PERP, TRIM25, VCL, SDC1, ARPC2, DLC1, CDH1
GOTERM_CC_FAT GO:0000346	Transcription Export Complex	0,025	THOC7, THOC2, THOC3
GOTERM_CC_FAT GO:0016323	Basolateral Plasma Membrane	0,039	TRIP6, NOTCH4, TRIM25, VCL, SDC1, ARPC2, DLC1, CDH1
KEGG_PATHWAY hsa05130	Pathogenic Escherichia coli infection	0,022	CD14, ARPC1B, ACTG1, ARPC2, CDH1

Apêndice 4: Tabela de genes dieferencialmente expressos comparando tumors primários e metástases pancreáticas

Probe ID	Gene Bank Acession	Probe Region	Gene Name	Gene ID	Gene Symbol	Ref Seq Gene	Chromossome	Log_2 (Fold Change M/T)
3712	BE763689	Exonic	death associated protein 3	7818	DAP3	NM_033657	chr1: 155691309 - 155695810	-7.67
605	AW938211	Exonic	thrombospondin 1	7057	THBS1	NM_003246	chr15: 39888085 - 39888654	-6.61
2497	BF888326	Exonic	palladin, cytoskeletal associated protein	23022	PALLD	NM_001166108	chr4: 169848516 - 169849101	-5.88
4073	AW805580	Exonic	Kruppel-like factor 10	7071	KLF10	NM_001032282	chr8: 103661173 - 103661729	-5.75
3234	BE840880	Exonic	serglycin	5552	SRGN	NM_002727	chr10: 70863894 - 70864339	-5.75
1324	BF923330	Exonic	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	3091	HIF1A	NM_181054	chr14: 62213753 - 62214259	-3.99
1376	BF378626	Exonic	guanine nucleotide binding protein (G protein), beta polypeptide 1	2782	GNB1	NM_002074	chr1: 1717243 - 1717778	-3.82
955	BE172273	Exonic	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked	1654	DDX3X	NM_001193416	chrX: 41208534 - 41209133	-3.42
3883	BE706846	Exonic	matrix-remodelling associated 7	439921	MXRA7	NM_001008528	chr17: 74676083 - 74676429	-3.29
4114	CK327212	Exonic	mitogen-activated protein kinase kinase 1	5604	MAP2K1	NM_002755	chr15: 66783397 - 66783674	-3.24
3702	BF357931	Exonic	Janus kinase 1	3716	JAK1	NM_002227	chr1: 65298979 - 65299505	-2.97
958	BE774819	Exonic	S100 calcium binding protein A8	6279	S100A8	NM_002964	chr1: 153362516 - 153362996	-2.90
4462	BE711811	Exonic	T-cell lymphoma invasion and metastasis 1	7074	TIAM1	NM_003253	chr21: 32492677 - 32493029	-2.66
3764	BQ328550	Exonic	nuclear receptor coactivator 4	8031	NCOA4	NM_001145260	chr10: 51589996 - 51590481	-2.58
2228	BF879985	Exonic	SMAD family member 2	4087	SMAD2	NM_005901	chr18: 45367901 - 45371790	-2.56
4111	BE833284	Exonic	LIM domain only 2 (rhombotin-like 1)	4005	LMO2	NM_001142315	chr11: 33880391 - 33880726	-2.49
3673	AW606595	Exonic	B-cell translocation gene 1, anti-proliferative	694	BTG1	NM_001731	chr12: 92536906 - 92537501	-2.37

1384	AW837526	Exonic	nuclear receptor coactivator 4	8031	NCOA4	NM_005437	chr10: 51589996 - 51590479	-2.33
3593	BF857286	Exonic	phosphatidylinositol binding clathrin assembly protein	8301	PICALM	NM_001008660	chr11: 85668821 - 85669229	-2.32
4305	BE086865	Exonic	leptin receptor overlapping transcript	54741	LEPROT	NM_001198681, NM_017526	chr1: 65898722 - 65899341	-2.31
4431	BE166669	Exonic	beta-2-microglobulin	567	B2M	NM_004048	chr15: 45003766 - 45010267	-2.29
2975	BF883034	Exonic	actin related protein 2/3 complex, subunit 1B, 41kDa; similar to Actin-related protein 2/3 complex subunit 1B (ARP2/3 complex 41 kDa subunit) (p41-ARC)	10095	ARPC1B	NM_005720	chr7: 98985809 - 98990358	-2.18
613	BF951773	Exonic	RAN, member RAS oncogene family	5901	RAN	NM_006325	chr12: 131357147 - 131360274	-2.07
2348	BF375449	Exonic	trans-golgi network protein 2	10618	TGOLN2	NM_006464	chr2: 85545357 - 85545693	-2.05
2972	BE720363	Exonic	insulin-like growth factor binding protein 7	3490	IGFBP7	NM_001553	chr4: 57898615 - 57907101	-2.05
992	AW606067	Exonic	cyclin-dependent kinase 2 associated protein 1	8099	CDK2AP1	NM_004642	chr12: 123745864 - 123746351	-2.00
1368	BE769759	Exonic	ubiquitin-conjugating enzyme E2E 3 (UBC4/5 homolog, yeast)	10477	UBE2E3	NM_182678	chr2: 181927700 - 181928032	-1.99
4117	AW602701	Exonic	RAB31, member RAS oncogene family	11031	RAB31	NM_006868	chr18: 9859318 - 9859483	-1.96
1362	AW994902	Exonic	ribosomal protein L15 pseudogene 22; ribosomal protein L15 pseudogene 18; ribosomal protein L15 pseudogene 17; ribosomal protein L15 pseudogene 3; ribosomal protein L15 pseudogene 7; ribosomal protein L15	6138	RPL15	NM_002948	chr3: 23961591 - 23962170	-1.96
4403	BE814543	Exonic	epithelial membrane protein 1	2012	EMP1	NM_001423	chr12: 13368770 - 13369416	-1.86
2756	AW939400	Exonic	transmembrane protein 50B	757	TMEM50B	NM_006134	chr21: 34821472 - 34821770	-1.84
2123	BQ349504	Exonic	lactate dehydrogenase A	3939	LDHA	NM_005566, NM_001135269, NM_001165414, NR_028500	chr11: 18427006 - 18429090	-1.82
2898	AW994882	Exonic	chromatin modifying protein 5	51510	CHMP5	NM_016410	chr9: 33281009 - 33281502	-1.78

4424	AW881287	Exonic	core-binding factor, beta subunit	865	CBFB	NM_022845	chr16: 67134484 - 67134831	-1.73
2116	AW860948	Exonic	SMAD family member 2	4087	SMAD2	NM_001003652	chr18: 45371731 - 45391505	-1.68
2500	BF090303	Exonic	B-cell CLL/lymphoma 6	604	BCL6	NM_001706	chr3: 187442740 - 187446240	-1.68
2499	BE693756	Exonic	cytoplasmic FMR1 interacting protein 1	23191	CYFIP1	NM_001033028	chr15: 23002937 - 23003558	-1.67
1710	BE774773	Exonic	amphiregulin; amphiregulin B	374	AREG	NM_001657	chr4: 75485911 - 75490396	-1.55
3625	BE706149	Exonic	eukaryotic translation initiation factor 3, subunit B	2512	FTL	NM_000146	chr19: 49469597 - 49470109	-1.48
4471	BF893510	Exonic	SIVA1, apoptosis-inducing factor	10572	SIVA1	NM_021709	chr14: 105221984 - 105225925	-1.40
2144	BF352147	Exonic	mitogen-activated protein kinase kinase kinase kinase 5	11183	MAP4K5	NM_198794	chr14: 50886314 - 50886798	-1.37
617	BF735925	Exonic	OCIA domain containing 1	54940	OCIAD1	NM_001079839	chr4: 48859228 - 48863139	-1.25
1729	BF377282	Exonic	serine incorporator 3	10955	SERINC3	NM_006811	chr20: 43128804 - 43132528	-1.18
2911	AW819770	Exonic	RAB4A, member RAS oncogene family	5867	RAB4A	NM_004578	chr1: 229434722 - 229439940	-1.15
2498	BE939301	Exonic	adrenergic, beta, receptor kinase 1	156	ADRBK1	NM_001619	chr11: 67046962 - 67049313	-1.13
1699	AW867343	Exonic	cullin 3	8452	CUL3	NM_003590	chr2: 225338728 - 225346721	-1.13
642	BF883122	Exonic	CD63 molecule	967	CD63	NM_001040034	chr12: 56119916 - 56121037	-0.95
666	BE814854	Exonic	chondroitin sulfate synthase 1	22856	CHSY1	NM_014918	chr15: 101716496 - 101716830	0.25
2595	BF945730	Exonic	protein tyrosine phosphatase, receptor type, B	5787	PTPRB	NM_001109754	chr12: 70915184 - 70928352	0.48
3919	BF082553	Exonic	RWD domain containing 4A	201965	RWDD4	NM_152682	chr4: 184560966 - 184561269	0.49
2711	AW369662	Exonic	prostate transmembrane protein, androgen induced 1	56937	PMEPA1	NM_199171	chr20: 56225817 - 56226437	0.49
2981	BE154053	Exonic	SMAD family member 5	4090	SMAD5	NM_001001419	chr5: 135496583 - 135508198	0.49
3677	AW862491	Exonic	TBC1 domain family, member 30	23329	TBC1D30	NM_015279	chr12: 65260591 - 65269059	0.50
1807	BE762926	Exonic	transcription factor 12	6938	TCF12	NM_003205	chr15: 57579444 - 57579689	0.53
2623	BE696994	Exonic	cell division cycle 2-like 5	8621	CDK13	NM_003718	chr7: 40027510 - 40037141	0.54

844	BE160080	Exonic	biorientation of chromosomes in cell division 1-like	259282	BOD1L	NM_148894	chr4: 13610162 - 13615197	0.55
2381	BE081005	Exonic	Guanine nucleotide binding protein (G protein), q polypeptide	2776	GNAQ	.	chr9: 80332512 - 80333071	0.56
4079	BF083029	Exonic	nuclear mitotic apparatus protein 1	4926	NUMA1	NM_006185	chr11: 71714302 - 71714553	0.56
2094	CK326980	Exonic	serine palmitoyltransferase, long chain base subunit 2	9517	SPTLC2	NM_004863	chr14: 77984398 - 78021760	0.56
1829	AW893012	Exonic	calcium/calmodulin-dependent protein kinase IV	814	CAMK4	NM_001744	chr5: 110819719 - 110820067	0.58
3606	BE168055	Exonic	LIM domain containing preferred translocation partner in lipoma	4026	LPP	NM_005578	chr3: 188605207 - 188605707	0.58
2466	BE840877	Exonic	protein kinase C, theta	5588	PRKCQ	NM_006257	chr10: 6498691 - 6521129	0.58
1608	BG001145	Exonic	nucleoporin 85kDa	79902	NUP85	NM_024844	chr17: 73224852 - 73225287	0.59
142	BE833257	Exonic	membrane associated guanylate kinase, WW and PDZ domain containing 1; CNKSR family member 3	154043	MAGI1	NM_015520	chr3: 65340767 - 65341124	0.59
4237	BE089334	Exonic	eukaryotic translation initiation factor 4E binding protein 2	1979	EIF4EBP2	NM_004096	chr10: 72185273 - 72185618	0.60
3720	AW889200	Exonic	similar to Down-regulated in metastasis protein (Key-1A6 protein) (Novel nucleolar protein 73) (NNP73); UTP20, small subunit (SSU) processome component, homolog (yeast)	27340	UTP20	NM_014503	chr12: 101779369 - 101779816	0.60
3162	AW606115	Exonic	endothelin converting enzyme 1	1889	ECE1	NM_001397	chr1: 21544367 - 21544858	0.62
4358	BE163446	Exonic	solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 23	79085	SLC25A23	NM_024103	chr19: 6440350 - 6440918	0.62
1397	BF803359	Exonic	MYC associated factor X	4149	MAX	NM_002382	chr14: 65543329 - 65560532	0.63
574	BF853165	Exonic	elaC homolog 2 (E. coli)	60528	ELAC2	NM_173717	chr17: 12908302 - 12915096	0.63
2962	AW749875	Exonic	spindlin 1	10927	SPIN1	NM_006717	chr9: 91091710 - 91092339	0.65
1370	BE702149	Exonic	protein kinase D3	23683	PRKD3	NM_005813	chr2: 37480321 - 37486784	0.65
3258	BG009603	Exonic	chromatin modifying protein 2A	27243	CHMP2A	NM_014453	chr19: 59063069 - 59065580	0.65

791	BE815309	Exonic	arrestin domain containing 4	91947	ARRDC4	NM_183376	chr15: 98516137 - 98516721	0.65
2617	BQ374821	Exonic	tudor domain containing 6	221400	TDRD6	NM_001168359	chr6: 46661368 - 46661784	0.65
3479	BE061111	Exonic	TAF8 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 43kDa	129685	TAF8	NM_138572	chr6: 42046925 - 42047215	0.66
2474	BF850918	Exonic	ADAM metallopeptidase with thrombospondin type 1 motif, 4	9507	ADAMTS4	NM_005099	chr1: 161163078 - 161163981	0.66
3682	BF881946	Exonic	integrin, alpha X (complement component 3 receptor 4 subunit)	3687	ITGAX	NM_000887	chr16: 31388528 - 31391643	0.67
4050	BF090332	Exonic	mitogen-activated protein kinase 3 (MAPK3), transcript variant 1, mRNA	5595	MAPK3	NM_001040056	chr16: 30128082 - 30129483	0.67
3373	BF947201	Exonic	neuron navigator 3; similar to neuron navigator 3	89795	NAV3	NM_014903	chr12: 78522513 - 78562594	0.68
1610	BG009879	Exonic	ST3 beta-galactoside alpha-2,3-sialyltransferase 1	6482	ST3GAL1	NM_173344	chr8: 134467539 - 134467974	0.69
2211	AW879352	Exonic	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 12 (CHL1-like helicase homolog, S. cerevisiae)	1663	DDX11	NM_004399	chr12: 31257454 - 31257719	0.69
1358	BE695748	Exonic	eukaryotic translation initiation factor 2-alpha kinase 2	5610	EIF2AK2	NM_001135651	chr2: 37336376 - 37347268	0.69
934	BF828830	Exonic	cullin-associated and neddylation-dissociated 2 (putative)	23066	CAND2	NM_001162499	chr3: 12875253 - 12875814	0.70
2007	BF805191	Exonic	nudix (nucleoside diphosphate linked moiety X)-type motif 16-like 1	84309	NUDT16L1	NM_001193452	chr16: 4745124 - 4745628	0.70
942	BQ377553	Exonic	FERM domain containing 4B	23150	FRMD4B	NM_015123	chr3: 69220050 - 69220338	0.71
1627	BE080712	Exonic	zinc finger, CCHC domain containing 6	79670	ZCCHC6	NM_024617	chr9: 88959973 - 88961395	0.72
2597	BF372924	Exonic	centrosomal protein 70kDa	80321	CEP70	NM_024491	chr3: 138224250 - 138248240	0.72
653	BF992143	Exonic	WAS/WASL interacting protein family, member 1	7456	WIPF1	NM_003387	chr2: 175431669 - 175431974	0.72
1348	AW880979	Exonic	TAP binding protein-like	55080	TAPBPL	NM_018009	chr12: 6562672 - 6566798	0.73
1827	BQ376493	Exonic	ribosomal protein S20	6224	RPS20	NM_001146227	chr8: 56985762 - 56986174	0.73

1407	BF899700	Exonic	zinc finger protein 652	22834	ZNF652	NM_001145365	chr17: 47375431 - 47375907	0.73
588	BE694447	Exonic	peptidylprolyl isomerase domain and WD repeat containing 1	23398	PPWD1	NM_015342	chr5: 64867758 - 64868113	0.76
3295	BF081836	Exonic	breast cancer metastasis suppressor 1	25855	BRMS1	NM_001024957	chr11: 66108770 - 66109650	0.77
3671	AW392280	Exonic	SECIS binding protein 2	79048	SECISBP2	NM_024077	chr9: 91972622 - 91973172	0.78
4382	BF769176	Exonic	RNA binding protein with multiple splicing	11030	RBPMS	NM_006867	chr8: 30423114 - 30423578	0.79
1352	BE089538	Exonic	chromosome 3 open reading frame 58	205428	C3orf58	NM_173552	chr3: 143710167 - 143710543	0.79
2565	AW602088	Exonic	DNA (cytosine-5-)-methyltransferase 1	1786	DNMT1	NM_001130823	chr19: 10246930 - 10248658	0.82
3343	BE833109	Exonic	S100 calcium binding protein A4	6275	S100A4	NM_002961	chr1: 153516194 - 153516400	0.82
1924	BF742190	Exonic	solute carrier family 38, member 1	81539	SLC38A1	NM_030674	chr12: 46578005 - 46578437	0.83
1612	BE160146	Exonic	tetratricopeptide repeat domain 23	64927	TTC23	NM_001040655	chr15: 99768787 - 99781744	0.84
4074	BF922401	Exonic	erythrocyte membrane protein band 4.1 (elliptocytosis 1, RH-linked)	2035	EPB41	NM_203343	chr1: 29391541 - 29424419	0.84
1316	BF829214	Exonic	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 2	8704	B4GALT2	NM_003780	chr1: 44450948 - 44456274	0.84
1798	BG006349	Exonic	Fc fragment of IgG, low affinity IIa, receptor (CD32)	2213	FCGR2B	NM_001002273	chr1: 161642782 - 161643863	0.85
2054	BE162994	Exonic	purine-rich element binding protein B	5814	PURB	NM_033224	chr7: 44919152 - 44919599	0.85
2227	BQ292121	Exonic	RAB3B, member RAS oncogene family	5865	RAB3B	NM_002867	chr1: 52384975 - 52385511	0.85
1735	AW361982	Exonic	nucleoporin 88kDa	4927	NUP88	NM_002532	chr17: 5289514 - 5292166	0.86
1603	AW804626	Exonic	ring finger protein 122	79845	RNF122	NM_024787	chr8: 33405996 - 33406385	0.86
2142	BE718669	Exonic	MAP/microtubule affinity-regulating kinase 2	2011	MARK2	NM_017490	chr11: 63672480 - 63676642	0.86
2766	BF882274	Exonic	chromosome 20 open reading frame 96	140680	C20orf96	NM_153269	chr20: 257949 - 270950	0.87
3367	BF910053	Exonic	mutL homolog 1, colon cancer,	4292	MLH1	NM_001167619	chr3: 37048511 - 37056018	0.87

2993	AW178990	Exonic	nonpolyposis type 2 (E. coli) nudix (nucleoside diphosphate linked moiety X)-type motif 1	4521	NUDT1	NM_002452	chr7: 2290462 - 2290702	0.88
1284	BG012009	Exonic	pleckstrin homology-like domain, family B, member 2; phosphatidylinositol-specific phospholipase C, X domain containing 2	257068	PHLDB2	NM_145753	chr3: 111630529 - 111630993	0.89
3334	BG005931	Exonic	breast cancer anti-estrogen resistance 3	8412	BCAR3	NM_003567	chr1: 94054875 - 94140324	0.89
1271	AW893496	Exonic	zinc finger protein 438	220929	ZNF438	NM_001143766	chr10: 31197860 - 31230582	0.90
1673	BF336884	Exonic	THO complex 3	84321	THOC3	NM_032361	chr5: 177309413 - 177310913	0.90
835	BF378914	Exonic	chromodomain helicase DNA binding protein 2	1106	CHD2	NM_001271	chr15: 93570117 - 93570515	0.90
3463	AW369635	Exonic	golgi SNAP receptor complex member 2	9570	GOSR2	NM_054022	chr17: 45017562 - 45017949	0.90
1248	BF857794	Exonic	zinc finger protein 585B	92285	ZNF585B	NM_152279	chr19: 37675462 - 37675895	0.90
1960	BF364914	Exonic	immunity-related GTPase family, Q	126298	IRGQ	NM_001007561	chr19: 44094912 - 44095268	0.91
3158	BG000783	Exonic	protease, serine, 12 (neurotrypsin, motopsin)	8492	PRSS12	NM_003619	chr4: 119201681 - 119202232	0.91
2444	AW883890	Exonic	UBA domain containing 2	337867	UBAC2	NM_001144072, NR_026644	chr13: 100038066 - 100038299	0.92
1778	BF087349	Exonic	mitogen-activated protein kinase kinase kinase 7 interacting protein 1	10454	TAB1	NM_153497	chr22: 39815559 - 39824176	0.92
784	BE828277	Exonic	THO complex 7 homolog (Drosophila)	80145	THOC7	NM_025075	chr3: 63819755 - 63822036	0.92
1468	AW753128	Exonic	purinergic receptor P2Y, G-protein coupled, 1	5028	P2RY1	NM_002563	chr3: 152554510 - 152554988	0.93
1047	BE163738	Exonic	similar to Ewing sarcoma breakpoint region 1; Ewing sarcoma breakpoint region 1	2130	EWSR1	NM_001163285	chr22: 29695791 - 29696442	0.93
2179	BE829286	Exonic	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	2065	ERBB3	NM_001982	chr12: 56495953 - 56496355	0.95
3348	BF956997	Exonic	insulin receptor	3643	INSR	NM_001079817	chr19: 7115432 - 7116001	0.95
922	BF926440	Exonic	chromosome 22 open reading	91689	C22orf32	NM_033318	chr22: 42477771 - 42478124	0.95

			frame 32					
1800	CK327215	Exonic	dihydropyrimidinase-like 2	1808	DPYSL2	NM_001386	chr8: 26492331 - 26505308	0.95
983	AW352153	Exonic	PRKC, apoptosis, WT1, regulator	5074	PAWR	NM_002583	chr12: 79986375 - 79990605	0.97
3945	BF357744	Exonic	WD repeat domain 89	112840	WDR89	NM_080666	chr14: 64065770 - 64066196	0.97
117	BF736083	Exonic	nephronophthisis 3 (adolescent); acyl-Coenzyme A dehydrogenase family, member 11	84129	ACAD11	NM_032169	chr3: 132277163 - 132277390	0.97
1759	AW799652	Exonic	tumor protein D52-like 2	7165	TPD52L2	NM_003288	chr20: 62522764 - 62522889	0.98
2196	AW862247	Exonic	DnaJ (Hsp40) homolog, subfamily C, member 9	23234	DNAJC9	NM_015190	chr10: 75002884 - 75003688	0.98
2160	BE832971	Exonic	phenylalanyl-tRNA synthetase, beta subunit	10056	FARSB	NM_005687	chr2: 223464681 - 223488446	0.99
520	BE168080	Exonic	family with sequence similarity 184, member A	79632	FAM184A	NM_024581	chr6: 119324732 - 119338111	0.99
1356	BE694456	Exonic	protein tyrosine phosphatase, receptor type, A	5786	PTPRA	NM_080841	chr20: 3001952 - 3005267	0.99
2917	BF952136	Exonic	TYRO3 protein tyrosine kinase	7301	TYRO3	NM_006293	chr15: 41865584 - 41870359	0.99
1046	BF813113	Exonic	zinc finger with KRAB and SCAN domains 5	23660	ZKSCAN5	NM_145102	chr7: 99130455 - 99130987	0.99
4095	BF330530	Exonic	von Hippel-Lindau tumor suppressor	7428	VHL	NM_000551	chr3: 10192836 - 10193261	0.99
3309	BQ317836	Exonic	homeobox A10	3206	HOXA10	NM_018951	chr7: 27211487 - 27219497	1.00
4354	BF997721	Exonic	tripartite motif-containing 25	7706	TRIM25	NM_005082	chr17: 54966236 - 54966713	1.02
4239	BF332449	Exonic	meningioma expressed antigen 5 (hyaluronidase)	10724	MGEA5	NM_001142434	chr10: 103554653 - 103555222	1.03
2193	BE833209	Exonic	N-myc (and STAT) interactor	9111	NMI	NM_004688	chr2: 152127313 - 152132079	1.04
4109	AW902244	Exonic	CaM kinase-like vesicle-associated	79012	CAMKV	NM_024046	chr3: 49895715 - 49896156	1.05
3119	BE156491	Exonic	major facilitator superfamily domain containing 8	256471	MFSD8	NM_152778	chr4: 128838980 - 128839490	1.06
3750	BE699157	Exonic	ataxin 7-like 3	56970	ATXN7L3	NM_020218	chr17: 42271767 - 42272857	1.06
3117	BE156258	Exonic	eukaryotic translation elongation factor 1 alpha-like 7; eukaryotic	645715	EEF1A1	NM_001402	chr6: 74225045 - 74225957	1.06

			translation elongation factor 1 alpha-like 3; similar to eukaryotic translation elongation factor 1 alpha 1; eukaryotic translation elongation factor 1 alpha 1					
1774	BF962290	Exonic	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)	2066	ERBB4	NM_005235	chr2: 212530050 - 212566894	1.06
1044	BF957176	Exonic	aryl hydrocarbon receptor nuclear translocator	405	ARNT	NM_178427	chr1: 150789833 - 150799060	1.06
1981	BF364095	Exonic	-	84131	CEP78	.	chr9: 80885760 - 80886427	1.07
99	BE181787	Exonic	transmembrane protein 33	55161	TMEM33	NM_018126	chr4: 41961229 - 41961552	1.08
3152	BF351473	Exonic	chromosome 1 open reading frame 115	79762	C1orf115	NM_024709	chr1: 220870418 - 220870743	1.08
3659	BQ315021	Exonic	vacuolar protein sorting 13 homolog A (S. cerevisiae)	23230	VPS13A	NM_001018037	chr9: 79999587 - 80000174	1.08
3994	BF928925	Exonic	leucine rich repeat containing 57	255252	LRRC57	NM_153260	chr15: 42836092 - 42839500	1.08
1454	AW389538	Exonic	RNA binding motif protein 19	9904	RBM19	NM_016196	chr12: 114387918 - 114395849	1.08
1928	BF809657	Exonic	fermitin family homolog 1 (Drosophila)	55612	FERMT1	NM_017671	chr20: 6088178 - 6093232	1.09
4380	BQ358748	Exonic	thioesterase superfamily member 5	284486	THEM5	NM_182578	chr1: 151823547 - 151826116	1.10
908	AW879619	Exonic	glyoxalase domain containing 4	51031	GLOD4	NM_016080	chr17: 673666 - 674684	1.10
3284	BF155304	Exonic	thymidine kinase 1, soluble	7083	TK1	NM_003258	chr17: 76170679 - 76171148	1.12
1812	BG005244	Exonic	RAB2A, member RAS oncogene family	5862	RAB2A	NM_002865	chr8: 61429727 - 61496846	1.14
2933	BF856859	Exonic	STE20-related kinase adaptor alpha	92335	STRADA	NM_001003786	chr17: 61780340 - 61780664	1.15
4116	BF956998	Exonic	Notch homolog 4 (Drosophila)	4855	NOTCH4	NM_004557	chr6: 32180251 - 32182014	1.16
1377	BF907266	Exonic	THO complex 2	57187	THOC2	NM_001081550	chrX: 122737000 - 122737309	1.16
3674	BE694434	Exonic	A kinase (PRKA) anchor protein 11	11215	AKAP11	NM_016248	chr13: 42877859 - 42891701	1.17
2822	BE163389	Exonic	cell division cycle and apoptosis regulator 1	55749	CCAR1	NM_018237	chr10: 70497598 - 70507224	1.17
789	AW995467	Exonic	solute carrier family 41, member 1	254428	SLC41A1	NM_173854	chr1: 205758657 - 205759152	1.18
526	BF836244	Exonic	fibronectin type III domain	64778	FNDC3B	NM_001135095	chr3: 171969147 -	1.18

			containing 3B				172025219	
3546	BF881000	Exonic	ATH1, acid trehalase-like 1 (yeast)	80162	ATHL1	NM_025092	chr11: 294616 - 294948	1.18
1037	BE003476	Exonic	ephrin-B2	1948	EFNB2	NM_004093	chr13: 107143416 - 107143809	1.19
2194	BQ367470	Exonic	colony stimulating factor 3 receptor (granulocyte)	1441	CSF3R	NM_000760	chr1: 36931736 - 36932164	1.20
1577	AW366182	Exonic	tetraspanin 18	90139	TSPAN18	NM_130783	chr11: 44952503 - 44952877	1.20
3315	BG876687	Exonic	nudE nuclear distribution gene E homolog 1 (A. nidulans)	4629	MYH11	NM_001040113, NM_001040114, NM_002474, NM_022844	chr16: 15814755 - 15818260	1.20
2482	BF924236	Exonic	prohibitin	5245	PHB	NM_002634	chr17: 47482389 - 47486445	1.20
649	BF754569	Exonic	proteasome maturation protein	51371	POMP	NM_015932	chr13: 29238644 - 29252216	1.21
2814	AW996919	Exonic	phosphoinositide-3-kinase, class 2, alpha polypeptide	5286	PIK3C2A	NM_002645	chr11: 17110151 - 17110608	1.21
3693	BG009501	Exonic	hect domain and RLD 2	8924	HERC2	NM_004667	chr15: 28359906 - 28362207	1.22
2935	BG000772	Exonic	CD14 molecule	929	CD14	NM_000591	chr5: 140011485 - 140011954	1.22
289	AW363728	Exonic	bystin-like	705	BYSL	NM_004053	chr6: 41900352 - 41900942	1.22
2183	BQ321689	Exonic	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7	5713	PSMD7	NM_002811	chr16: 74334032 - 74339324	1.22
2866	BF854041	Exonic	transforming, acidic coiled-coil containing protein 2	10579	TACC2	NM_006997	chr10: 124008155 - 124013595	1.23
1458	AW880788	Exonic	LIM domain kinase 1	3984	LIMK1	NM_002314	chr7: 73523237 - 73530187	1.23
3280	BF334504	Exonic	neurofibromin 2 (merlin)	4771	NF2	NM_000268	chr22: 30070859 - 30077552	1.24
4432	BE925852	Exonic	RAB13, member RAS oncogene family; similar to hCG24991	5872	RAB13	NM_002870	chr1: 153954403 - 153954626	1.25
943	BE694208	Exonic	inhibitor of growth family, member 1	3621	ING1	NM_005537	chr13: 111372152 - 111372457	1.26
961	BF924125	Exonic	myeloid/lymphoid or *mix*ed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 10	8028	MLLT10	NM_004641	chr10: 21959410 - 21962598	1.27
2945	BE813923	Exonic	AT rich interactive domain 4A	5926	ARID4A	NM_023001	chr14: 58838692 - 58839233	1.27

			(RBP1-like)					
3651	AW176028	Exonic	dual specificity phosphatase 4	1846	DUSP4	NM_001394	chr8: 29194220 - 29194574	1.27
1615	BE926631	Exonic	ankyrin repeat domain 40	91369	ANKRD40	NM_052855	chr17: 48772971 - 48773355	1.27
2870	BF920836	Exonic	transcription factor binding to IGHM enhancer 3	7030	TFE3	NM_006521	chrX: 48886720 - 48887130	1.28
1353	BQ348843	Exonic	leucine aminopeptidase 3	51056	LAP3	NM_015907	chr4: 17583937 - 17597115	1.31
2533	BE174220	Exonic	cell division cycle 16 homolog (S. cerevisiae)	8881	CDC16	NM_003903	chr13: 115037779 - 115038054	1.32
4078	BG015476	Exonic	interferon, alpha-inducible protein 6	2537	IFI6	NM_002038	chr1: 27992916 - 27994805	1.32
3369	BF805352	Exonic	general transcription factor IIIC, polypeptide 5, 63kDa	9328	GTF3C5	NM_012087	chr9: 135933417 - 135933681	1.37
2097	CK326976	Exonic	heat shock protein 90kDa alpha (cytosolic), class A member 1; heat shock protein 90kDa alpha (cytosolic), class A member 1	3320	HSP90AA1	NM_001017963	chr14: 102549925 - 102551156	1.40
2925	CK327113	Exonic	C-terminal binding protein 1	1487	CTBP1	NM_001328	chr4: 1209812 - 1232003	1.41
849	BF378856	Exonic	SH3 and PX domains 2A	9644	SH3PXD2A	NM_014631	chr10: 105355438 - 105355991	1.42
1776	AW380133	Exonic	integrin, beta 7	3695	ITGB7	NM_000889	chr12: 53585158 - 53585802	1.43
2159	BG009760	Exonic	T-cell leukemia translocation altered gene	6988	TCTA	NM_022171	chr3: 49453065 - 49453401	1.43
1306	BF353211	Exonic	chromodomain helicase DNA binding protein 2	1106	CHD2	NM_001271	chr15: 93569622 - 93572586	1.44
2580	BF956989	Exonic	A kinase (PRKA) anchor protein 2; paralemmin 2/ PALM2-AKAP2 readthrough transcript	11217/445815	AKAP2/ PALM2-AKAP2	NM_001004065, NM_001136562, NM_001198656/NM_007203, NM_147150	chr9: 112900635 - 112900980	1.49
1668	BQ338380	Exonic	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	84901	NFATC2IP	NM_032815	chr16: 28976015 - 28976343	1.53
3964	AW748714	Exonic	zinc finger and SCAN domain containing 18	65982	ZSCAN18	NM_023926	chr19: 58599594 - 58600025	1.56

2098	CK326979	Exonic	replication factor C (activator 1) 4, 37kDa (RFC4), transcript variant 3	5984	RFC4	NM_002916	chr3: 186508123 - 186510680	1.59
4143	BQ302049	Exonic	heat shock protein 90kDa beta (Grp94), member 1	7184	HSP90B1	NM_003299	chr12: 104325336 - 104328065	1.60
2053	BF880524	Exonic	poly(A) binding protein, cytoplasmic 1-like	80336	PABPC1L	NM_001124756	chr20: 43545394 - 43547919	1.60
4302	BF882251	Exonic	tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)	7035	TFPI	NM_006287	chr2: 188329444 - 188329970	1.62
3381	BE172434	Exonic	protein kinase C and casein kinase substrate in neurons 2	11252	PACSIN2	NM_007229	chr22: 43266963 - 43267277	1.67
2564	BG003639	Exonic	neural cell adhesion molecule 1	4684	NCAM1	NM_000615	chr11: 113076801 - 113078666	1.72
4036	BF154406	Exonic	lysine (K)-specific demethylase 6A	7403	KDM6A	NM_021140	chrX: 44970789 - 44971059	1.81
2310	AW860708	Exonic	heterogeneous nuclear ribonucleoprotein U-like 2	221092	HNRNPUL2	NM_001079559	chr11: 62481869 - 62482255	2.00
2932	BE063167	Exonic	gametogenetin binding protein 2	79893	GGNBP2	NM_024835	chr17: 34935789 - 34937908	2.08
993	AW748486	Exonic	StAR-related lipid transfer (START) domain containing 3	10948	STARD3	NM_001165937	chr17: 37819074 - 37819526	2.37
941	BE172353	Exonic	far upstream element (FUSE) binding protein 1	8880	FUBP1	NM_003902	chr1: 78414317 - 78425940	2.41
2052	AW380113	Exonic	G patch domain containing 4	54865	GPATCH4	NM_182679	chr1: 156564541 - 156564956	2.52
2118	AW804490	Exonic	WW domain binding protein 11	51729	WBP11	NM_016312	chr12: 14943467 - 14946857	2.92
4386	BE926343	Exonic	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	4313	MMP2	NM_004530	chr16: 55519535 - 55525803	2.93
3670	BE089406	Exonic	oxysterol binding protein-like 3	26031	OSBPL3	NM_145322	chr7: 24843932 - 24846456	3.03
1671	BF753293	Exonic	chromosome 19 open reading frame 21	126353	C19orf21	NM_173481	chr19: 759907 - 763637	3.09
1670	AW902314	Exonic/ ncRNA	family with sequence similarity 36, member A/ non-protein coding RNA 201 (NCRNA00201)	116228/ 284702	FAM36A/ NCRNA00201	NM_198076/ NR_266778	chr1: 245007377 - 245007734	1.67
2337	BE836530	Intergenic	-	-	-	.	chr8: 124510217 - 124510551	-2.77
1661	BF894831	Intergenic	-	-	-	.	chr9: 114696678 - 114697175	-1.91

1296	BF768642	Intergenic	-	-	-	.	chrX: 139842199 - 139842531	0.31
83	CK327083	Intergenic	-	-	-	.	chr12: 106742160 - 106742749	0.50
2405	BE710975	Intergenic	-	-	-	.	chr3: 141333071 - 141333471	0.55
4280	BF944704	Intergenic	-	-	-	.	chr1: 11353583 - 11355287	0.57
436	BF364063	Intergenic	-	-	-	.	chr17: 17148059 - 17148369	0.59
543	BF992202	Intergenic	-	-	-	.	chr14: 21671137 - 21671545	0.65
3226	BF928916	Intergenic	-	-	-	.	chrX: 48437775 - 48438272	0.65
3974	AW902458	Intergenic	-	-	-	.	chr1: 214213913 - 214214349	0.76
3900	AW819863	Intergenic	-	-	-	.	chr17: 62409845 - 62410200	0.87
809	BE141719	Intergenic	-	-	-	.	chr2: 207182963 - 207183584	0.90
3897	BE155946	Intergenic	-	-	-	.	chr20: 1025153 - 1025834	0.91
426	BF944705	Intergenic	-	-	-	.	chr7: 99637117 - 99637572	0.95
513	AW937867	Intergenic	-	-	-	.	chr2: 176496659 - 176497171	0.97
4268	BF918966	Intergenic	-	-	-	.	chr5: 92193518 - 92194056	0.98
3596	BF837556	Intergenic	-	-	-	.	chr13: 43460581 - 43460994	1.00
1018	BF956245	Intergenic	-	-	-	.	chr7: 99345995 - 99346500	1.01
2848	BF736784	Intergenic	-	-	-	.	chr1: 35644977 - 35645390	1.06
909	AW936043	Intergenic	-	-	-	.	chr11: 77371632 - 77372107	1.12
3602	BG012678	Intergenic	-	-	-	.	chr5: 173251870 - 173252290	1.12

987	AW391850	Intergenic	-	-	-	.	chr14: 23020736 - 23021031	1.19
3977	BF754591	Intergenic	-	-	-	.	chr1: 46653707 - 46654108	1.38
3499	BF363994	Intergenic	-	-	-	.	chr17: 45152177 - 45152608	1.70
3865	BE008094	Intergenic	-	-	-	.	chr2: 181710096 - 181710451	1.79
4262	BE702744	Intergenic	-	-	-	.	chr7: 127944519 - 127944846	1.81
2450	BF893217	Intergenic	-	-	-	.	chr20: 3910164 - 3910595	1.84
1973	AW835362	Intronic	Insulin-like growth factor binding protein 7	3490	IGFBP7	NM_001553	chr4: 57928550 - 57929060	-8.49
3494	BF364433	Intronic	Sequestosome 1	8878	SQSTM1	NM_001142298	chr5: 179254631 - 179254940	-0.28
2785	CK327138	Intronic	Prefoldin subunit 1	5201	PFDN1	NM_002622	chr5: 139633755 - 139634141	0.44
1662	BE069131	Intronic	Phospholipase C, beta 1 (phosphoinositide-specific)	23236	PLCB1	NM_182734	chr20: 8638142 - 8638719	0.47
3887	BE156492	Intronic	A kinase (PRKA) anchor protein 2	11217	AKAP2	NM_001004065, NM_001198656	chr9: 112815008 - 112815506	0.51
2009	BF803619	Intronic	Poliovirus receptor-related 2 (herpesvirus entry mediator B)	5819	PVRL2	NM_001042724	chr19: 45379546 - 45379991	0.53
846	BE160464	Intronic	Gephyrin	10243	GPHN	NM_001024218	chr14: 66990664 - 66991047	0.54
1297	CK327188	Intronic	Thyroid hormone receptor, beta (erythroblastic leukemia viral (v-erb-a) oncogene homolog 2, avian)	7068	THRB	NM_001128177	chr3: 24295120 - 24295537	0.55
2448	CK327209	Intronic	Dystrophin (muscular dystrophy, Duchenne and Becker types)	1756	DMD	NM_000109	chrX: 33002957 - 33003298	0.56
2456	BF365245	Intronic	Centaurin, gamma 2	116987	AGAP1	NM_001037131	chr2: 236669614 - 236670139	0.60
3509	AW835433	Intronic	Vesicle transport through interaction with t-SNAREs homolog 1A (yeast)	143187	VTI1A	NM_145206	chr10: 114468638 - 114468946	0.63
446	BF919473	Intronic	Catenin (cadherin-associated protein), delta 1	1500	CTNND1	NM_001085458	chr11: 57541563 - 57541976	0.67
2789	BF897833	Intronic	WD repeat domain 37	22884	WDR37	NM_014023	chr10: 1140522 - 1141062	0.67

3920	BE174039	Intronic	Ras association (RalGDS/AF-6) and pleckstrin homology domains 1	65059	RAPH1	NM_213589	chr2: 204351542 - 204352046	0.67
2821	CK327191	Intronic	Protein phosphatase 3 (formerly 2B), catalytic subunit, beta isoform	5532	PPP3CB	NM_021132	chr10: 75242804 - 75243166	0.69
3873	BF371714	Intronic	N-myc downstream regulated gene 1	10397	NDRG1	NM_001135242	chr8: 134284139 - 134284576	0.70
3522	BF335132	Intronic	PDGFA associated protein 1	11333	PDAP1	NM_014891	chr7: 98997093 - 98997637	0.71
3124	AW880867	Intronic	Chromosome 2 open reading frame 34	79823	C2orf34	NM_024766	chr2: 44873009 - 44873447	0.71
1283	CK327189	Intronic	Mediator complex subunit 12-like	116931	MED12L	NM_053002	chr3: 150952711 - 150953165	0.72
1936	CK327033	Intronic	alanine-glyoxylate aminotransferase 2-like 2	85007	AGXT2L2	NM_153373	chr5: 177657886 - 177658326	0.74
544	BF736673	Intronic	Chromosome 10 open reading frame 26	54838	C10orf26	NM_017787	chr10: 104537616 - 104538000	0.74
2064	BF736301	Intronic	RAD54-like (S. cerevisiae)	8438	RAD54L	NM_001142548	chr1: 46738550 - 46738858	0.74
1581	BE156252	Intronic	Sine oculis binding protein homolog (Drosophila)	55084	SOBP	NM_018013	chr6: 107919688 - 107920245	0.76
3229	AW937483	Intronic	Activating signal cointegrator 1 complex subunit 3	10973	ASCC3	NM_006828	chr6: 101180631 - 101181205	0.76
1609	AW748493	Intronic	RNA binding motif protein 47	54502	RBM47	NM_001098634	chr4: 40563872 - 40564416	0.79
1298	BF992848	Intronic	Death-associated protein kinase 1	1612	DAPK1	NM_004938	chr9: 90121444 - 90121900	0.80
1676	AW879689	Intronic	Ring finger protein 43	54894	RNF43	NM_017763	chr17: 56433010 - 56433176	0.81
1293	BF368584	Intronic	DnaJ (Hsp40) homolog, subfamily C, member 3	5611	DNAJC3	NM_006260	chr13: 96432041 - 96432369	0.84
3886	AW820168	Intronic	Syntaxin 12	23673	STX12	NM_177424	chr1: 28141725 - 28142251	0.84
790	AW860958	Intronic	Dystonin	667	DST	NM_015548	chr6: 56329341 - 56329893	0.84
3881	BE141765	Intronic	NHL repeat containing 2	374354	NHLRC2	NM_198514	chr10: 115632894 - 115633410	0.85
2784	BF857841	Intronic	Inositol 1,4,5-triphosphate receptor, type 1	3708	ITPR1	NM_001099952	chr3: 4786881 - 4787338	0.86
842	BG009898	Intronic	Signal peptide peptidase 3	121665	SPPL3	NM_139015	chr12: 121212026 - 121212458	0.86
2435	AW84667	Intronic	Chromosome 11 open reading	79096	C11orf49	NM_001003676	chr11: 47169433 - 47169741	0.88

	0		frame 49					
1575	BF998295	Intronic	Solute carrier family 43, member 2	124935	SLC43A2	NM_152346	chr17: 1505024 - 1505501	0.88
3912	BG004239	Intronic	neuroblastoma breakpoint family, member 10	100132406	NBPF10	NM_001039703	chr1: 145302128 - 145302563	0.89
1629	CK327094	Intronic	Activating transcription factor 2	1386	ATF2	NM_001880	chr2: 175949486 - 175949865	0.89
3175	AW836835	Intronic	Zinc finger protein 659	79750	ZNF385D	NM_024697	chr3: 21467925 - 21468460	0.89
3202	AW837982	Intronic	Abl interactor 2	10152	ABI2	NM_005759	chr2: 204250129 - 204250461	0.90
1682	BF893233	Intronic	FIG4 homolog (S. cerevisiae)	9896	FIG4	NM_014845	chr6: 110120316 - 110120726	0.92
3100	BF846679	Intronic	Interleukin 4 receptor	3566	IL4R	NM_000418	chr16: 27332628 - 27333150	0.95
2434	AW837930	Intronic	KIAA1217	56243	KIAA1217	NM_001098500	chr10: 24782052 - 24782506	0.98
1631	BE081324	Intronic	Mitogen-activated protein kinase 1	5594	MAPK1	NM_002745	chr22: 22160497 - 22160963	0.99
1681	AW749983	Intronic	Usher syndrome 1C (autosomal recessive, severe)	10083	USH1C	NM_153676	chr11: 17524178 - 17524643	0.99
2037	BF896571	Intronic	Pleckstrin homology, Sec7 and coiled-coil domains 1(cytohesin 1)	9267	CYTH1	NM_017456	chr17: 76679340 - 76679807	0.99
4352	BE003587	Intronic	Protein phosphatase 4, regulatory subunit 1	9989	PPP4R1	NM_001042388	chr18: 9561281 - 9561760	0.99
76	BE160056	Intronic	Na+/K+ transporting ATPase interacting 2	154215	NKAIN2	NM_001040214	chr6: 124980949 - 124981492	1.06
1199	BF364145	Intronic	Solute carrier family 4, sodium bicarbonate cotransporter, member 7	9497	SLC4A7	NM_003615	chr3: 27466211 - 27466598	1.07
1643	BF894801	Intronic	Mitogen-activated protein kinase kinase kinase 1	4214	MAP3K1	NM_005921	chr5: 56125509 - 56125997	1.08
3847	BF749179	Intronic	stonin 2	85439	STON2	.	chr14: 81732820 - 81733207	1.08
2758	BF848956	Intronic	Myosin VA (heavy chain 12, myoxin)	4644	MYO5A	NM_001142495	chr15: 52660950 - 52661377	1.11
4064	BQ316935	Intronic	Phosphonoformate immuno-associated protein 5	10443	N4BP2L2	NM_033111	chr13: 33109399 - 33109744	1.13
157	BF348899	Intronic	Dehydrogenase/reductase (SDR family) X-linked	207063	DHRSX	NM_145177	chrY: 2324676 - 2324945	1.15
3903	BE166370	Intronic	Nuclear receptor subfamily 2, group C, member 2	7182	NR2C2	NM_003298	chr3: 15084054 - 15084337	1.17

1639	AW836810	Intronic	Phosphatase and tensin homolog (mutated in multiple advanced cancers 1)	5728	PTEN	NM_000314	chr10: 89630175 - 89630699	1.23
3842	BF334142	Intronic	Enolase superfamily member 1	55556	ENOSF1	NM_001126123	chr18: 709089 - 709430	1.25
525	BF368582	Intronic	Kelch domain containing 4	54758	KLHDC4	NM_017566	chr16: 87774429 - 87774951	1.28
912	AW865910	Intronic	Nucleolar protein NOP5/NOP58	51602	NOP58	NM_015934	chr2: 203164022 - 203164376	1.28
484	AW805690	Intronic	Heterogeneous nuclear ribonucleoprotein L	3191	HNRNPL	NM_001533	chr19: 39339100 - 39339450	1.28
829	BE156374	Intronic	Calponin 3, acidic	1266	CNN3	NM_001839	chr1: 95374991 - 95375338	1.33
4339	BF898694	Intronic	Megakaryoblastic leukemia (translocation) 1	57591	MKL1	NM_020831	chr22: 40849141 - 40849644	1.36
2436	BE700386	Intronic	Chromosome 12 open reading frame 65	91574	C12orf65	NM_152269	chr12: 123728281 - 123728616	1.37
3513	AW901803	Intronic	Gliomedin	342035	GLDN	NM_181789	chr15: 51642970 - 51643317	1.38
1584	CK327079	Intronic	T-cell lymphoma invasion and metastasis 2	26230	TIAM2	.	chr6: 155366476 - 155366837	1.38
1641	BF357735	Intronic	Vesicle transport through interaction with t-SNAREs homolog 1A (yeast)	143187	VTI1A	NM_145206	chr10: 114432448 - 114432669	1.38
2369	AW834040	Intronic	RNA binding motif protein 47	54502	RBM47	NM_001098634	chr4: 40464007 - 40464480	1.39
1200	AW880165	Intronic	Transforming growth factor, beta receptor II (70/80kDa)	7048	TGFBR2	NM_001024847	chr3: 30723134 - 30723319	1.39
1194	BF944729	Intronic	Arrestin, beta 1	408	ARRB1	NM_020251	chr11: 75045505 - 75045972	1.40
2411	BF894929	Intronic	Ubiquitin specific peptidase 15	9958	USP15	NM_006313	chr12: 62725137 - 62725661	1.42
1690	BF926453	Intronic	X-ray repair complementing defective repair in Chinese hamster cells 1	7515	XRCC1	NM_006297	chr19: 44076178 - 44076504	1.43
875	BF894811	Intronic	Tyrosine kinase 2	7297	TYK2	NM_003331	chr19: 10474182 - 10474612	1.44
3946	BF328128	Intronic	Ewing tumor-associated antigen 1	54465	ETAA1	NM_019002	chr2: 67624934 - 67625337	1.47
3185	BE181783	Intronic	Casein kinase 1, delta	1453	CSNK1D	NM_001893	chr17: 80226176 - 80226555	1.49
3230	AW865762	Intronic	Rap guanine nucleotide exchange factor (GEF) 2	9693	RAPGEF2	NM_014247	chr4: 160215173 - 160215501	1.49
2384	BE168993	Intronic	Macrophage erythroblast attacher	10296	MAEA	NM_001017405	chr4: 1318234 - 1318651	1.54
3223	BF754634	Intronic	Mitogen-activated protein kinase kinase kinase 14	9020	MAP3K14	NM_003954	chr17: 43384670 - 43385102	1.55

105	CK327134	Intronic	CAP-GLY domain containing linker protein family, member 4	79745	CLIP4	NM_024692	chr2: 29382944 - 29383185	1.55
1951	BF332457	Intronic	protein kinase C and casein kinase substrate in neurons 2	11252	PACSIN2	NM_001184970	chr22: 43311439 - 43311997	1.57
4319	CK327106	Intronic	Tyrosine kinase, non-receptor, 2	10188	TNK2	NM_005781	chr3: 195591793 - 195592189	1.58
4379	BF368581	Intronic	Sorbin and SH3 domain containing 1	10580	SORBS1	NM_015385	chr10: 97315461 - 97315809	1.60
1305	BF857201	Intronic	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	10257	ABCC4	NM_005845	chr13: 95848163 - 95848548	1.60
1230	BF881469	Intronic	Alstrom syndrome 1	7840	ALMS1	NM_015120	chr2: 73772958 - 73773485	1.65
4258	BF364131	Intronic	Chromosome 16 open reading frame 46	123775	C16orf46	NM_152337	chr16: 81104689 - 81105025	1.70
136	AW902654	Intronic	Ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)	5168	ENPP2	NM_001040092	chr8: 120624467 - 120624862	1.74
2459	AW935931	Intronic	SEC31 homolog A (S. cerevisiae)	22872	SEC31A	NM_016211	chr4: 83784014 - 83784389	1.75
4253	BE089575	Intronic	Chromosome 20 open reading frame 7	79133	C20orf7	NM_001039375	chr20: 13789721 - 13790060	1.75
3979	CK327186	Intronic	Leptin receptor	3953	LEPR	NM_002303	chr1: 66071979 - 66072278	1.76
1168	AW852151	Intronic	DIP2 disco-interacting protein 2 homolog B (Drosophila)	57609	DIP2B	NM_173602	chr12: 51007113 - 51007544	1.79
3207	BF336807	Intronic	Leucine rich repeat containing 16	55604	LRRC16A	NM_017640	chr6: 25437044 - 25437384	1.88
3097	BE008080	Intronic	OCIA domain containing 1	54940	OCIAD1	NM_017830	chr4: 48837837 - 48838145	1.89
858	BG010309	Intronic	DAZ associated protein 1	26528	DAZAP1	NM_170711	chr19: 1419237 - 1419769	1.89
3144	BG004232	Intronic	PHD finger protein 12	57649	PHF12	NM_020889	chr17: 27268945 - 27269327	1.90
2051	BF875190	Intronic	Bardet-Biedl syndrome 5	129880	BBS5	NM_152384	chr2: 170352422 - 170352845	1.91
4360	BE168353	Intronic	PAS domain containing serine/threonine kinase	23178	PASK	NM_015148	chr2: 242086059 - 242086526	1.99
2347	BE156164	Intronic	zinc finger protein 586	54807	ZNF586	NM_017652	chr19: 58288803 - 58289127	2.23
3849	BE007885	Intronic	Rho GTPase activating protein 12	94134	ARHGAP12	NM_018287	chr10: 32168422 - 32168792	2.25
3585	BF875569	Intronic	Activating transcription factor 3	467	ATF3	NM_001030287	chr1: 212785642 - 212786212	2.39

814	AW819948	Intronic	Protein phosphatase 2 (formerly 2A), regulatory subunit B, alpha isoform	5520	PPP2R2A	NM_002717	chr8: 26194381 - 26194828	2.46
914	BF893298	Intronic	Mitogen-activated protein kinase kinase 5	5607	MAP2K5	NM_002757	chr15: 67889806 - 67890177	2.79
3483	BE062809	Intronic/ Intronic	TMEM189-UBE2V1 readthrough/ ubiquitin-conjugating enzyme E2 variant 1	387522/ 7335	TMEM189-UBE2V1/ UBE2V1	NM_001032288, NM_021988, NM_022442, NM_199144/NM_199203	chr20: 48718697 - 48719099	1.36
4307	BE087324	Intronic/ Intronic	microcephalin 1 (MCPH1), transcript variant 1/ Angiopoietin 2	79648/ 285	MCPH1/ ANGPT2	NM_024596/ NM_001118887, NM_001118888, NM_001147	chr8: 6382399 - 6382774	1.94
432	AW880028	ncRNA	Homo sapiens non-protein coding RNA 85 (NCRNA00085), non-coding RNA	147650	NCRNA00085	NR_024330	chr19: 52199048 - 52199487	0.74
3986	BF893393	ncRNA	Homo sapiens small nucleolar RNA, C/D box 89 (SNORD89), small nucleolar RNA	692205	SNORD89	NR_003070	chr2: 101889292 - 101889671	0.78
2345	AW366194	ncRNA	Homo sapiens hypothetical LOC387647 (LOC387647), non-coding RNA	387647	LOC387647	NR_003930	chr10: 29710515 - 29711100	0.86
1648	AW601510	ncRNA	Chromosome 3 open reading frame 47	339942	C3orf47	NR_026991	chr3: 129036816 - 129037420	0.93
2833	BF875772	ncRNA	Homo sapiens myocardial infarction associated transcript (non-protein coding) (MIAT), transcript variant 1, non-coding RNA	440823	MIAT	NR_033321	chr22: 27069703 - 27070091	0.98
2438	BF364175	ncRNA	Homo sapiens hypothetical LOC92973 (LOC92973), non-coding RNA	92973	LOC92973	NR_024006	chr9: 35861748 - 35862051	1.27

Apêndice 5: Análise de enriquecimento (DAVID) de genes diferencialmente expressos na assinatura de amostras de tumores primários comparados à metástase pancreática.

Termos	Categorias	p-valor*	Genes
codificador de proteína			
BIOCARTA h_hifPathway	Hypoxia-Inducible Factor in the Cardiovascular System	0.011	ARNT, LDHA, HSP90AA1, HIF1A, VHL
GOTERM_BP_FAT GO:0006403	RNA localization	0.026	NUP88, THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0051236	establishment of RNA localization	0.017	NUP88, THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0015931	nucleobase, nucleoside, nucleotide and nucleic acid transport	0.026	NUP88, THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0050658	RNA transport	0.017	NUP88, THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0050657	nucleic acid transport	0.017	NUP88, THOC7, THOC2, RAN, NUP85, THOC3
GOTERM_BP_FAT GO:0051168	nuclear export	0.019	THOC7, STRADA, THOC2, RAN, THOC3
GOTERM_BP_FAT GO:0006405	RNA export from nucleus	0.038	THOC7, THOC2, RAN, THOC3
GOTERM_BP_FAT GO:0019048	virus-host interaction	0.038	THOC7, THOC2, INSR, THOC3
GOTERM_CC_FAT GO:0031982	vesicle	0.009	RAB13, PRSS12, CAMKV, RAB2A, ECE1, RAN, WIPF1, TMEM33, PICALM, SRGN, TGOLN2, HSP90AA1, GGNBP2, MYH11, RAB4A, PACSIN2, RAB3B, HSP90B1, PIK3C2A, THBS1
GOTERM_CC_FAT GO:0031410	cytoplasmic vesicle	0.005	RAB13, PRSS12, CAMKV, RAB2A, ECE1, RAN, WIPF1, TMEM33, PICALM, SRGN, TGOLN2, HSP90AA1, GGNBP2, MYH11, RAB4A, PACSIN2, RAB3B, HSP90B1, PIK3C2A, THBS1
GOTERM_CC_FAT GO:0016023	cytoplasmic membrane-bounded vesicle	0.015	RAB2A, ECE1, RAN, WIPF1, TMEM33, PICALM, SRGN, TGOLN2, HSP90AA1, GGNBP2, MYH11, RAB4A, PACSIN2, RAB3B, PIK3C2A, HSP90B1, THBS1
GOTERM_CC_FAT GO:0031988	membrane-bounded vesicle	0.022	RAB2A, ECE1, RAN, WIPF1, TMEM33, PICALM, SRGN, TGOLN2, HSP90AA1, GGNBP2, MYH11, RAB4A, PACSIN2, RAB3B, PIK3C2A, HSP90B1, THBS1
***locus* gênico de transcritos intrônicos**			
GOTERM_CC_FAT GO:0043228	non-membrane-bounded organelle	0.016	ABI2, GPHN, ITPR1, SORBS1, TYK2, NDRG1, DST, MCPH1, USH1C, MAEA, BBS5, SLC4A7, RAPH1, CNN3, NR2C2, DMD, DAZAP1, MAP3K1, MAP2K5, PHF12, NOP5/NOP58, ATF3, ALMS1, STON2, DAPK1, MYO5A, MAPK1

GOTERM_CC_FAT GO:0043232	intracellular non-membrane-bounded organelle	0.016	ABI2, GPHN, ITPR1, SORBS1, TYK2, NDRG1, DST, MCPH1, USH1C, MAEA, BBS5, SLC4A7, RAPH1, CNN3, NR2C2, DMD, DAZAP1, MAP3K1, MAP2K5, PHF12, NOP5/NOP58, ATF3, ALMS1, STON2, DAPK1, MYO5A, MAPK1
KEGG_PATHWAY hsa04010	MAPK signaling pathway	0.021	ARRB1, MAP3K1, ATF2, MAP3K14, MAP2K5, TGFBR2, PPP3CB, RAPGEF2, MAPK1
GOTERM_BP_FAT GO:0006793	phosphorus metabolic process	0.042	ABI2, ENPP2, PTEN, CSNK1D, MAP3K1, MAP2K5, TYK2, PPP3CB, PPP2R2A, PASK, MAP3K14, TNK2, TGFBR2, DAPK1, MAPK1
GOTERM_BP_FAT GO:0006796	phosphate metabolic process	0.042	ABI2, ENPP2, PTEN, CSNK1D, MAP3K1, MAP2K5, TYK2, PPP3CB, PPP2R2A, PASK, MAP3K14, TNK2, TGFBR2, DAPK1, MAPK1
GOTERM_CC_FAT GO:0005856	cytoskeleton	0.042	ABI2, DMD, GPHN, MAP3K1, ITPR1, SORBS1, MAP2K5, TYK2, NDRG1, MCPH1, DST, MAEA, ALMS1, BBS5, RAPH1, DAPK1, MYO5A, CNN3, MAPK1
GOTERM_BP_FAT GO:0030029	actin filament-based process	0.048	ABI2, SORBS1, PACSIN2, CNN3, MYO5A, DST

Apêndice 6: Tabela de genes associados com comportamento metastático ou cancer em outros tipos de câncer

Gene	Tipo de tecido ou câncer	Descrição	Referência
Correlacionados com comportamento metastático			
Taxa de sobrevida			
CYFIP1	Câncer colorectal	Baixa taxa de sobrevida em câncer colorectal	(Silva et al. 2009)
Expressão em metástase			
NOTCH4	Carcinoma saliva adenóide cístico	Expressão foi significativamente maior em pancientes que possuíam metástase	(Ding et al.)
LDHA	Carcinoma cervical	Expressão diminuída em carcinoma cervical com metástase no linfonodo	(Huang et al.)
SH3PXD2A	Câncer de próstata	Aumento da expressão promovida por Runx2 e está correlacionado com comportamento metastático	(Baniwal et al.)
TFPI	mama	TFPI foi detectado em células cancerosas no linfonodo metastático	(Sierko et al.)
Prognóstico			
CBFB	Células epiteliais	Mau prognóstico em células tumorais epiteliais.	(Mendoza-Villanueva et al.)
WIPF1	Colorectal, mama, glioma	Baixa expressão de genes co-expressos com WIPF1 apresenta bom prognóstico	(Staub et al. 2009)
STARD3	Câncer de mama	Níveis elevados em pacientes apresenta mau prognóstico	(Ou et al. 2009)
Resistência ao tratamento			
MARK2	Câncer cervical	HeLa resistentes à cisplatina apresentam maior expressão de MARK2.	(Wu et al.)
Crescimento independente de ancoragem			
CCAR1	Linhagens de câncer de cólon	Depleção de CCAR1 suprime o crescimento independente de ancoragem.	(Ou et al. 2009)
Correlação com comportamento tumoral			
Desregularão em outros cânceres			
KDM6A	Células epiteliais	HPV16 E7 proteína induz a produção de KDM6A	(McLaughlin-Drubin et al.)
RAB2A	Câncer vaginal	Proteína está aumentada em câncer	(Lomnytska et al.)
Presentes em metástases com alteração invertida observada no câncer de pâncreas			
MHY11	Mama	Diminuído em metástase de mama	(Seitz et al. 2006)
PAWR	Mama	Diminuição de PAWR está associado com mau prognóstico no câncer de pâncreas	(Nagai et al.)
PRKCQ	Fibroblasto	PKC teta como potencial supressor do programa de invasão	(Stapleton et al. 2002)
DPYSL2	Câncer colorectal	Diminuição está associada com metástase no câncer colorectal	(Nakagawa et al. 2004)

ANEXOS ELETRÔNICOS

- **Trabalhos publicados**

TAHIRA, A. C., Kubrusly, Márcia S, Faria, Michele F, Dazzani, Bianca, Fonseca, Rogério S, Maracaja-Coutinho, Vinicius, Verjovski-Almeida, Sergio, Machado, Marcel CC, Reis, Eduardo M. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. Molecular Cancer. v.10, p.141, 2011.

ANEXO

SÚMULA CURRICULAR

DADOS PESSOAIS

Nome: Ana Carolina Tahira

Nascimento: 01 de outubro de 1984

Naturalidade: São Paulo/SP

Nacionalidade: Brasileira

Endereço: Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748 sala 1200, 05508-000 São Paulo, SP. Fone: (011) 3091-2173 r. 207.

e-mail: actahira@iq.usp.br, tahira.ana@gmail.com

FORMAÇÃO ACADÊMICA:

- Química – Bacharelado em Química no Instituto de Química da Universidade de São Paulo, sito à Av. Prof. Lineu Prestes 748 – Cidade Universitária – Butantã – CEP: 05508-000 – São Paulo/SP. Período: 2003 a 2006.
- Doutorado em Ciências Biológicas (Bioquímica). Instituto de Química da Universidade de São Paulo, sito à Av. Prof. Lineu Prestes 748 – Cidade Universitária – Butantã – CEP: 05508-000 – São Paulo/SP. Período: 2006 até o presente.

HISTÓRICO PROFISSIONAL:

03/2004 - 03/2005 Aluna de iniciação cientfífica no Instituto de Pesquisa Energéticas e Nucleares (IPEN) – Cidade Universitária – Butantã.

Projeto: Validação do método de análise por ativação com nêutrons para determinação de antimônio (Sb) em amostras ambientais.

Orientador: Mitiko Saiki

Bolsista CNPq

05/2005 – 12/2006 Aluna de inciciação científica no Departamento de Bioquímica, Insituto de Química – Universidade de São Paulo.

Projeto: Caracterização de transcrito intrônicos identificado como diferencialmente expresso em carcinoma de célula renal.

Orientador: Sergio Verjovski-Almeida

Bolsista CNPq

02/2007 – Atual Aluna de Doutorado Direto, Departamento de Bioquímica, Instituto de Química – Universidade de São Paulo

Projeto: Identificação de marcadores moleculares associados ao câncer de pâncreas utilizando *microarrays* de DNA.

Orientador: Eduardo Moraes R. Reis

Bolsa CNPq: Vigência da bolsa: 05/2007 a 09/2007

Bolsa FAPESP: Vigência da Bolsa: 01/10/2007 a 31/05/2012

PUBLICAÇÕES

Artigos completos publicados em periódicos

TAHIRA, A. C., Kubrusly, Márcia S, Faria, Michele F, Dazzani, Bianca, Fonseca, Rogério S, Maracaja-Coutinho, Vinicius, Verjovski-Almeida, Sergio, Machado, Marcel CC, Reis, Eduardo M. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. Molecular Cancer. v.10, p.141, 2011.

Trabalhos apresentados na forma oral em congressos nacionais

TAHIRA, A. C., KUBRUSLY, M., FARIA, M. F., VERJOVSKI-ALMEIDA, S., MACHADO, M. C. C., REIS, E. M. Identificação de assinaturas de expressão gênica correlacionadas à transformação maligna e metástase no câncer de pâncreas utilizando microarranjos de DNA enriquecidos em RNAs não-codificadores, 2008. 56º CBG (Congresso Brasileiro de Genética) – Prêmio pós-graduação em Genética e Evolução Humana e Genética Médica.

Resumos Publicados em anais de Congresso

AYUPE-OLIVEIRA, A. C., **TAHIRA, A. C.**, VERJOVSKI-ALMEIDA, S., REIS, E. M. REIS. Biogenesis, stability and sub-cellular localization of long non-coding RNAs expressed in intronic regions of the human genome. In: 58º Congresso Brasileiro de Genética, 2012, Foz do Iguaçu – PR.

AYUPE-OLIVEIRA, A. C., **TAHIRA, A. C.**, VERJOVSKI-ALMEIDA, S., REIS, E. M. REIS. Biogenesis, stability and sub-cellular localization of long non-coding RNAs expressed in intronic regions of the human genome. In: Regulatory & Non-coding RNAs, 2012, p.119.

FACHEL, AA ; **TAHIRA, A. C.** ; VILELLA-ARIAS, Santiago A. ; MARACAJA-COUTINHO, V. ; GIMBA, E. ; Vignal, G ; REIS, E. M. R. ; Verjovski-Almeida, S. . Long ncRNAs expressed in renal cell carcinoma are correlated to malignancy and survival: functional associations. In: 72nd Harden Conference - RNA regulators of gene expression, 2012, Cambridge. Abstracts 72nd Harden Conference - RNA regulators of gene expression, 2012. p. 33-33.

FACHEL A ; **TAHIRA, A. C**. ; VILELLA-ARIAS, S. A. ; MARACAJA-COUTINHO, V. ; GIMBA, E. ; Vignal, G ; REIS, E. M. R. ; VERJOVSKI-ALMEIDA, S. . Long ncRNAs expressed in renal cell carcinoma and correlated to malignancy or to survival. In: Functional Genomics and Systems Biology Conference 2011, 2011, Cambridge. Abstracts Functional Genomics & Systems Biology Conference 2011, 2011.

TAHIRA, A. C. ; KUBRUSLY, M. ; FARIA, M. F. ; DAZZANI, B. ; VERJOVSKI-ALMEIDA, S. ; MACHADO, M. C. C. ; REIS, E. M. . Signatures of long intronic non-coding RNAs differentially expressed in primary tumors and metastatic cancer. In: The Non-Coding Genome, 2010, Heidelberg. The Non-Coding Genome, 2010.

FACHEL, A. A. ; **TAHIRA, A. C.** ; Maracaja-Coutinho, Vinicius ; GIMBA, E. ; VIGNAL, G.M. ; LOURO, R. ; REIS, E. M. ; VERJOVSKI-ALMEIDA, S. . RNAs não-codificadores intrônicos longos correlacionados com carcinogênese e sobrevida em câncer de rim. In: 56º Congresso Brasileiro de Genética, 2010, Guarujá - SP. 56º Congresso Brasileiro de Genética, 2010.

FACHEL, A. A. ; **TAHIRA, A. C.** ; Maracaja-Coutinho, Vinicius ; GIMBA, E. ; CAMPOS, F. ; LOURO, R. ; REIS, E. M. ; VERJOVSKI-ALMEIDA, S. . Long intronic noncoding RNAs signatures of malignancy and survival outcome in clear cell carcinoma. In: Systems biology: global regulation of gene expression, 2010, New York. Systems biology: global regulation of gene expression, 2010.

TAHIRA, A. C. ; KUBRUSLY, M. ; FARIA, M. F. ; VERJOVSKI-ALMEIDA, S. ; REIS, E. M. ; MACHADO, M. C. C. . Gene Expression Profiling reveals Long Intronic non-coding RNAs differentially Expresssed in Pancreatic cancer and Metastasis. In: APA- forty-first annual Meeting, 2010, Chicago, Illinois. APA- forty-first annual Meeting, 2010.

TAHIRA, A. C. ; KUBRUSLY, M. ; FARIA, M. F. ; VERJOVSKI-ALMEIDA, S. ; MACHADO, M. C. C. ; REIS, E. M. . Long intronic noncoding RNAs expression signatures correlated to metastasis in pancreatic cancer. In: XXXIX Annual Meeting of SBBq, 2010, Foz do Iguaçu, Paraná. XXXIX Annual Meeting of SBBq, 2010.

TAHIRA, A. C. ; KUBRUSLY, M. ; FARIA, M. F. ; VERJOVSKI-ALMEIDA, S. ; REIS, E. M. R. ; MACHADO, M. C. C. . An intronic ncRNA expression signature of metastases in pancreatic cancer. In: AACR - American Association for Cancer Research, 2009. AACR 100th Annual Meeting 2009, 2009.

FACHEL, A. A. ; **TAHIRA, A. C.** ; MOREIRA, Y. B. ; GIMBA, E. ; CAMPOS, F. ; LOURO, R. ; REIS, E. M. ; VERJOVSKI-ALMEIDA, S. . Intronic noncoding RNAs correlated with carcinogenesis and survival in clear cell renal cell carcinoma. In: AACR - 100th Annual Meeting 2009, 2009, Denver, Colorado. 100th - AACR Annual Meeting 2009, 2009.

TAHIRA, A. C. ; KUBRUSLY, M. ; FARIA, M. F. ; VERJOVSKI-ALMEIDA, S. ; MACHADO, M. C. C. ; REIS, E. M. R. . Identifcação de assinaturas de expressão gênica correlacionadas à transformação maligna e metástase no câncer de pâncreas utilizando microarranjos de DNA enriquecidos em RNAs não-codifcadores. In: 54º Congresso Brasileiro de Genética, 2008, Salvador. 54º Congresso Brasileiro de genética, 2008.

TAHIRA, A. C. ; KUBRUSLY, M. ; FARIA, M. F. ; CUNHA,J.E.M. ; JUKEMURA,J ; BACCHELLA,T. ; VERJOVSKI-ALMEIDA, S. ; REIS, E. M. ; MACHADO, M. C. C. . IDENTIFICATION OF PROTEIN-CODING AND NONCODING GENE EXPRESSION SIGNATURES CORRELATED WITH MALIGNANT TRANSFORMATION AND METASTASES IN PANCREATIC CANCER.. In: 39th Annual APA Meeting, 2008, Chicago. 39th Annual APA Meeting, 2008.

FACHEL, A. A. ; **TAHIRA, A. C.** ; MOREIRA, Y. B. ; JUNDI, T. E. ; GIMBA, E. ; VIGNAL, G.M. ; CAMPOS, F. ; LOURO, R. ; REIS, E. M. ; VERJOVSKI-ALMEIDA, S. . Intronic noncoding RNAs correlated to carcinogenesis and patient survival in clear cell Renal Cell Carcinoma. In: 3rd EFS Functional Genomics Conference, 2008, Innsbruck. 3rd EFS Functional Genomics Conference, 2008.

FARIA, M. F. ; **TAHIRA, A. C.** ; KUBRUSLY, M. ; MACHADO, M. C. C. ; VERJOVSKI-ALMEIDA, S. ; REIS, E. M. . Identificação de RNAs não codificadores intrônicos diferencialmente expressos entre amostras de câncer de pâncreas primário e metastático. In: 16º SIICUSP, 2008, Ribeirão Preto, SP. 16º SIICUSP, 2008.

TAHIRA, A. C. ; FACHEL, A. A. ; LOURO, R. ; REIS, E. M. R. ; VERJOVSKI-ALMEIDA, S. . Differential expression of intronic noncoding RNAs in renal cell carcinoma using quantitative RT-PCR. In: Sociedade Brasileira de Bioquímica, 2007, Salvador. XXXVI Reunião Anual da SBBq, 2007, Salvador - BA, 2007.

TAHIRA, A. C. ; KUBRUSLY, M. ; FONSECA, R. ; VERJOVSKI-ALMEIDA, S. ; MACHADO, M. C. C. ; REIS, E. M. . Identificação de marcadores moleculares associados ao câncer de pâncreas utilizando microarranjos de DNA enriquecidos em RNAs intrônicos não-

codificantes. In: 53º Congresso Brasileiro de Genética, 2007, Águas de Lindóia, SP. Resumos do 53º Congresso Brasileiro de Genética, 2007.

TAHIRA, A. C. ; FACHEL, A. A. ; LOURO, R. ; VERJOVSKI-ALMEIDA, S. . Caracterização de transcritos intrônicos diferencialmente expressos em carcinoma de células renais por microarrays e PCR fita específica. In: 14º SIICUSP, 2006, Ribeirão Preto, SP. 14º SIICUSP, 2006.

FACHEL, A. A. ; EL-JUNDI, T. A. ; **TAHIRA, A. C.** ; CAMPOS, F. ; REIS, E. M. ; VERJOVSKI-ALMEIDA, S. . Differential expression of intronic non-coding RNAs in Renal Cell carcinoma using cDNA Microarray analysis. In: Mechanisms&MOdels of Cancer, 2006, Cold Spring Harbor, NY. Mechanisms & Models of Cancer - Cold Spring Harbor Conference, 2006.

TAHIRA, A. C. ; FACHEL, A. A. ; LOURO, R. ; VERJOVSKI-ALMEIDA, S. . Caracterização de transcritos intrônicos diferencialmente expressos em carcinoma de células renais por análise de real-time. In: XXIII - Semana da Química, 2006, São Paulo, SP. XXIII - Semana da Química, 2006.

ATIVIDADES EXTRAS

Orientação da aluna de graduação em Química (IQ-USP/São Paulo) Michele Felix Faria durante Iniciação Científica no laboratório do Prof. Eduardo Moraes Reis e sob sua supervisão.

> Projeto: Validação da expressão de transcritos intrônicos não-codificadores, através de PCR quantitativo em tempo real, em amostras de tecido de adenocarcinoma de pâncreas primário e metastático.
>
> Bolsa CNPq: Vigência da Bolsa: 2008-06/2009

Orientação da aluna de graduação em Biologia (IB-USP/São Paulo) Bianca Dazzani durante Iniciação Científica no laboratório do Prof. Eduardo Moraes Reis e sob sua supervisão.

> Projeto: Caracterização funcional do ncRNA intrônico PPP3CB super-expresso em metástases de tumores pancreáticos utilizando ensaios de apoptose.
>
> Bolsa FAPESP: Vigência da Bolsa: 01/02/2010 a 31/1/2012

Monitoria nos Cursos:

2007 – II Curso de Inverno “Temas de Bioquímica e Biologia Molecular” (Palestras sobre *Microarray* e expressão gênica). Universidade de São Paulo, USP, Insituto de Química São Paulo, Brasil.

2008 – III Curso de Inverno "Temas de Bioquímica e Biologia Molecular" (Palestras sobre *Microarray* e expressão gênica). Universidade de São Paulo, USP, Insituto de Química São Paulo, Brasil.

2010 – V Curso de Inverno "Temas de Bioquímica e Biologia Molecular" (Palestras sobre *Microarray* e expressão gênica). Universidade de São Paulo, USP, Insituto de Química São Paulo, Brasil.

Cursos de extensão:

2012 – 20hs V Curso de Bioinformática: Algoritmos e técnicas computacionais para montagem e análise de genomas. LacTaD, UNICAMP, Campinas, São Paulo, Brasil.

2011 – 38hs Curso de Verão em Biofísica Molecular. Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, São José do Rio Preto, Sao Paulo, Brasil.

2004 – 15hs As Cores como o químico as vê.Instituto de Química - USP, IQ-USP, Brasil.

2004 – 20hs Continuing Education in Proteção radiológica para radioquímica. Instituto de Pesquisas Energéticas e Nucleares, IPEN, Sao Paulo, Brasil.

2004 – 8hs Continuing Education in Farmacologia Quântica. Universidade Federal de São Carlos, UFSCAR, Sao Carlos, Brasil.