



UNIVERSIDADE FEDERAL DO PARANÁ

BRANCA ENGEL TIMONER

ESTUDO DA EXPRESSÃO DE LNCRNAs CONTENDO SEQUÊNCIAS DE
ELEMENTOS TRANSPONÍVEIS EM CÂNCER

CURITIBA

2021

UNIVERSIDADE FEDERAL DO PARANÁ

BRANCA ENGEL TIMONER

ESTUDO DA EXPRESSÃO DE LNCRNAs CONTENDO SEQUÊNCIAS DE
ELEMENTOS TRANSPONÍVEIS EM CÂNCER

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Genética do Departamento de Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, como requisito parcial para obtenção do título de Mestre em Genética.

Orientadora: Prof^ª. Dr^ª. Jaqueline Carvalho de Oliveira.

Co-orientador: Dr. Leandro Encarnação Garcia.

CURITIBA

2021

Universidade Federal do Paraná. Sistema de Bibliotecas.
Biblioteca de Ciências Biológicas.
(Rosilei Vilas Boas – CRB/9-939).

Timoner, Branca Engel.

Estudo da expressão de lncRNAs contendo sequências de elementos transponíveis no câncer. / Branca Engel Timoner. – Curitiba, 2021.
64 f. : il.

Orientadora: Jaqueline Carvalho de Oliveira.
Coorientador: Leandro Encarnação Garcia.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Ácido ribonucleico. 2. Testes biológicos. 3. Marcadores biológicos. 4. Câncer - Pesquisa. 5. Biomarcadores. I. Título. II. Oliveira, Jaqueline Carvalho de. III. Garcia, Leandro Encarnação. IV. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

CDD (20. ed.) 616.994



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS BIOLÓGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA -
40001016006P1

ATA Nº336

ATA DE SESSÃO PÚBLICA DE DEFESA DE Mestrado PARA A OBTENÇÃO DO GRAU DE MESTRE EM GENÉTICA

No dia vinte e cinco de março de dois mil e vinte e um às 15:00 horas, na sala Teams, Departamento de Genética, foram instaladas as atividades pertinentes ao rito de defesa de dissertação da mestranda BRANCA ENGEL TIMONER, intitulada: **Estudo da Expressão de lncRNAs Contendo Sequências de Elementos Transponíveis no Câncer**, sob orientação da Profa. Dra. JAQUELINE CARVALHO DE OLIVEIRA. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: JAQUELINE CARVALHO DE OLIVEIRA (UNIVERSIDADE FEDERAL DO PARANÁ), DANIEL PACHECO BRUSCHI (UNIVERSIDADE FEDERAL DO PARANÁ), DALILA LUCIOLA ZANETTE (FUNDAÇÃO OSWALDO CRUZ). A presidência iniciou os ritos definidos pelo Colegiado do Programa e, após exarados os pareceres dos membros do comitê examinador e da respectiva contra argumentação, ocorreu a leitura do parecer final da banca examinadora, que decidiu pela APROVAÇÃO. Este resultado deverá ser homologado pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais definidos pelo programa. A outorga de título de mestre está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, JAQUELINE CARVALHO DE OLIVEIRA, lavrei a presente ata, que vai assinada por mim e pelos demais membros da Comissão Examinadora.

CURITIBA, 25 de Março de 2021.

Assinatura Eletrônica

26/03/2021 10:32:28.0

JAQUELINE CARVALHO DE OLIVEIRA

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

26/03/2021 15:36:51.0

DANIEL PACHECO BRUSCHI

Avallador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

29/03/2021 10:53:46.0

DALILA LUCIOLA ZANETTE

Avallador Externo (FUNDAÇÃO OSWALDO CRUZ)

Setor de Ciências Biológicas, Centro Politécnico - CURITIBA - Paraná - Brasil
CEP 81531-980 - Tel: (41) 3361-1587 - E-mail: ppg-gen@ufpr.br

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.

Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 85184

Para autenticar este documento/assinatura, acesse <https://www.prrpg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp>
e insira o código 85184



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS BIOLÓGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA -
40001016006P1

ATA Nº336

**MINUTES OF PUBLIC DEFENSE SESSION OF THESIS TO OBTAIN THE DEGREE
OF MASTER OF SCIENCES IN GENÉTICA**

On 25/03/2021 at 15:00, In Teams, Departamento de Genética of the Federal University of Paraná, the activities pertaining to the rite of defense of work entitled: *Estudo da Expressão de LncRNAs Contendo Sequências de Elementos Transponíveis no Câncer* by BRANCA ENGEL TIMONER, under the supervision of Dr. JAQUELINE CARVALHO DE OLIVEIRA, were undertaken. The Examining Board, designated by the Faculty of the Graduate Program of the Federal University of Paraná in GENÉTICA was composed of the following Members: JAQUELINE CARVALHO DE OLIVEIRA (UNIVERSIDADE FEDERAL DO PARANÁ), DANIEL PACHECO BRUSCHI (UNIVERSIDADE FEDERAL DO PARANÁ), DALILA LUCIOLA ZANETTE (FUNDAÇÃO OSWALDO CRUZ). The Board Chair initiated the rites as defined by the Faculty of the Program and, after the remarks of the members of the examining committee and the respective responses were addressed, the final evaluation of the examining board was read, which was decided by APPROVAL. This result should be made official by the Faculty of the Program, through the fulfillment of all recommendation and corrections requested by the examining board within the procedural deadlines defined by the program. The granting of the title of master of sciences is contingent upon the fulfillment of all the requirements and terms determined in the regulations of the Graduate Program. There being no further business to transact, the Chair adjourned the session, of which, I, JAQUELINE CARVALHO DE OLIVEIRA, drew up these minutes, which are signed by me and the other members of the Examining Board.

CURITIBA, March 25th, 2021.

Electronic Signature

26/03/2021 10:32:28.0

JAQUELINE CARVALHO DE OLIVEIRA

President of the Examining Board (UNIVERSIDADE FEDERAL DO PARANÁ)

Electronic Signature

26/03/2021 15:36:51.0

DANIEL PACHECO BRUSCHI

Internal Member (UNIVERSIDADE FEDERAL DO PARANÁ)

Electronic Signature

29/03/2021 10:53:46.0

DALILA LUCIOLA ZANETTE

External Member (FUNDAÇÃO OSWALDO CRUZ)

AGRADECIMENTOS

À minha orientadora Prof^a Dra. Jaqueline Carvalho de Oliveira pela excelente orientação, pela disposição e paciência, por me tranquilizar e me compreender em momentos de dificuldades, por tudo que me ensinou nesses 2 anos de orientação e pela amizade.

Aos membros da banca examinadora, Prof. Dr. Daniel Pacheco Bruschi e Dra. Dalila Luciola Zanette por aceitarem analisar este trabalho.

À Dra. Adriana Ludwig pela eficiente colaboração neste trabalho e por todos os ensinamentos da área da escrita de um artigo e da área do mundo dos transposons.

À quase Dra. Carolina Mathias e à Dra. Érika Pereira Zambalde pela sempre disponibilidade em colaborar, pelos ensinamentos, pela motivação, por serem um exemplo de cientista e pela amizade.

Ao meu co-orientador Dr. Leandro Encarnação Garcia pelos ensinamentos e colaboração.

À Prof^a Dra. Enilze Maria de Souza Fonseca Ribeiro obrigada por toda a ajuda no desenvolvimento e correção deste trabalho.

À Prof^a Dra. Daniela Fiori Gradia por todas as discussões e ajuda no desenvolvimento do projeto.

Aos colegas do Laboratório de Citogenética Humana e Oncogenética por toda a ajuda, companheirismo e paciência.

Agradeço àquelas que além de colegas de mestrado se tornaram amigas – Luana Kamarowski Alves e Cristiane Sato Mara Muller.

Ao meu esposo Leo por desde o começo me incentivar a fazer o mestrado e pelo apoio durante esses dois anos.

Às agências de fomento CNPq, Fundação Araucária e a CAPES, que possibilitaram o desenvolvimento desse trabalho.

RESUMO

As células cancerosas frequentemente acumulam múltiplas mutações e anormalidades na expressão gênica, incluindo em longos RNAs não codificantes (lncRNAs). A maioria dos lncRNAs, entretanto, ainda é pouco explorada. Recentemente, foi proposto que os elementos transponíveis (TEs) podem atuar como domínios funcionais de lncRNA. Nesta hipótese, fragmentos de TE são cooptados para formar RIDLs (*Repeat Insertion Domains of LncRNAs*). Nós investigamos lncRNAs diferencialmente expressos em oito tipos de câncer em comparação com suas contrapartes não tumorais, usando dados do Atlas do Genoma do Câncer. Dentre eles, selecionamos lncRNAs contendo sequências de RIDLs (lncRNA-RIDLs). Encontramos 944 genes lncRNA-RIDLs expressos diferencialmente no câncer. Três lncRNA-RIDLs se destacaram em pelo menos sete tipos de câncer: SNHG3, MAGI2-AS3 e PVT1. Nos 20 genes mais diferencialmente expressos de lncRNA-RIDLs, 44% foram associados ao tempo de sobrevivência em pelo menos um tipo de tumor; os relacionados à sobrevivência na maioria dos tipos de câncer foram CARMN, LINC00261, LINC00511 e MIR99AHG. Além disso, entre uma compilação de 122 lncRNAs com papéis causais no fenótipo do câncer, 19 (15,6%) tinham RIDLs. Nosso estudo contribui para a investigação dos genes lncRNA-RIDLs e indica novas possibilidades para a identificação de biomarcadores no câncer.

Palavras-chave: lncRNAs, RIDL, elementos transponíveis, expressão diferencial, marcador molecular, câncer

ABSTRACT

Cancer cells often accumulate multiple mutations and gene expression abnormalities, including in long non-coding RNAs (lncRNAs). Most lncRNAs, however, are still underexplored. It has been recently proposed that transposable elements (TEs) can act as functional lncRNA domains. In this hypothesis, TE fragments are co-opted to form RIDLs (Repeat Insertion Domains of LncRNAs). We investigated lncRNAs differentially expressed in eight types of cancer compared to their non-tumoral counterparts, using data from The Cancer Genome Atlas. Among them, we selected lncRNAs containing sequences of RIDLs (lncRNA-RIDLs). We found 944 differentially expressed lncRNA-RIDLs genes in cancer. Three lncRNA-RIDLs stood out in at least seven types of cancer: SNHG3, MAGI2-AS3, and PVT1. In the top 20 most differentially expressed lncRNA-RIDLs genes, 44% were associated with survival time in at least one tumor type; the ones related to survival in most cancer types were CARMN, LINC00261, LINC00511, and MIR99AHG. Additionally, among a compilation of 122 lncRNAs with causal roles in cancer phenotype, 19 (15.6%) had RIDLs. Our study contributes to the investigation of lncRNA-RIDLs genes and indicates new possibilities for the identification of biomarkers in cancer.

Keywords: lncRNAs, RIDL, transposable elements, differential expression, molecular marker, cancer.

LISTA DE ILUSTRAÇÕES

REVISÃO DE LITERATURA

Figura 1 - LncRNAs associados à carcinogênese	16
Figura 2 - Classificação dos lncRNAs em relação a sua posição com a de genes codificantes de proteínas próximos	17
Figura 3 - Mecanismos de ação de lncRNAs	18
Figura 4 - Os diversos mecanismos de mobilização do transposon	21
Figura 5 - Exemplos de perfis de inserção de DNAs repetitivos em lncRNAs.....	23
Figura 6 - Modelo RIDL	24

CAPÍTULO I

Figure 1 - Schematic representation of the identification of lncRNA-RIDLs genes differentially expressed in the eight types of cancer analyzed in this work	36
Figure 2 - Top 20 lncRNA-RIDLs genes differentially expressed	42

LISTA DE TABELAS

CAPÍTULO I

Table 1 - Number of lncRNA genes differentially expressed in eight cancer types, their respective intersections with the lncRNA-RIDLs, and the number of exclusive lncRNA-RIDLs for each cancer type	36
Table 2 - Number of lncRNA-RIDLs genes that were downregulated (below and left in the diagonal) and upregulated (above and right in the diagonal) in the eight cancer types and the pairwise comparisons of common lncRNA-RIDLs genes between types of cancers	37
Table 3 - Top 20 differentially expressed lncRNA-RIDLs genes in each analyzed cancer type.....	38
Table 4 - lncRNA-RIDLs genes associated with survival time in cancer	41
Table 5 - lncRNA-RIDLs among the compilation of lncRNAs with causal roles in cancer phenotype	44

SUMÁRIO

1	INTRODUÇÃO	11
2	REVISÃO DE LITERATURA	14
2.1	CÂNCER	14
2.2	LONGOS RNAS NÃO CODIFICANTES	15
2.3	ELEMENTOS TRANSPONÍVIES	19
2.4	ELEMENTOS RIDLS	23
3	OBJETIVOS	26
3.1	OBJETIVO GERAL	26
3.2	OBJETIVOS ESPECÍFICOS	26
4	MATERIAL E MÉTODOS	27
4.1	EXTRAÇÃO DE DADOS E ANÁLISE DE EXPRESSÃO DIFERENCIAL	27
4.2	IDENTIFICAÇÃO DOS GENES LNCRNA-RIDLS EXPRESSOS DIFERENCIALMENTE EM CÂNCERES	27
4.3	ANÁLISE DE SOBREVIVÊNCIA	28
4.4	IDENTIFICAÇÃO DE GENES LNCRNA-RIDLS COM UM PAPEL SIGNIFICATIVO NO CÂNCER	28
5	CAPÍTULO I	29
6	CONCLUSÃO	60
	REFERÊNCIAS	61

1 INTRODUÇÃO

O câncer é a primeira ou a segunda causa de morte prematura em 134 dos 183 países em todo o mundo, com mais de 18,1 milhões de novos casos e 9,6 milhões de mortes por ano (Sung et al. (2021)). O desenvolvimento desta doença envolve a aquisição de algumas vantagens pelas células, como evasão da supressão do crescimento, evasão do sistema imunológico, sinalização proliferativa sustentável, imortalidade replicativa, indução de angiogênese, invasão e metástase, desregulação do metabolismo celular, promoção da inflamação, instabilidade genômica e modificações epigenéticas (Hanahan; Weinberg (2011)).

Recentemente, as taxas de sobrevida na maioria dos tipos de tumor aumentaram, em parte devido à implementação de marcadores moleculares de diagnóstico e prognóstico. Embora a maioria dos marcadores moleculares seja baseada em regiões codificantes do genoma, um número crescente de estudos mostra também a importância das regiões não codificantes, como RNAs não codificantes (Oliveira et al. (2019); Anastasiadou et al. (2019)).

Os rápidos avanços tecnológicos no sequenciamento de alto rendimento revelaram muitos milhares de novas classes de RNAs não codificantes, e estes incluem os longos RNAs não codificantes (lncRNAs). lncRNAs são transcritos com mais de 200 nucleotídeos de comprimento, e mais frequentemente transcritos pela RNA polimerase II, podem ter íntrons e variantes de splicing, podem exibir cauda poli A na extremidade 3' e podem sofrer modificações epigenéticas semelhantes a genes que codificam proteínas, como como regulação pela metilação do DNA e mudanças nas histonas (Yao et al. (2019)). Os lncRNAs podem estar presentes tanto no núcleo quanto no citoplasma e participar da regulação da expressão gênica em diferentes níveis (transcricional, pós-transcricional e pós-translacional) (Geisler; Coller (2013); Fang; Fullwood (2016)).

Dezenas de milhares de genes de lncRNAs foram catalogados (Hon et al. (2019)) até agora, mas a maioria deles é mal compreendida. A diferenciação de genes lncRNA reais de ruído transcricional e a caracterização de seus domínios funcionais são úteis para a identificação mais rápida de lncRNAs com papel em processos

fisiológicos e doenças. Pouco se sabe sobre a origem desses genes; no entanto, elementos transponíveis (TEs) têm sido associados ao surgimento e diversificação de lncRNAs (Hadjigryrou; Delihias (2013); Kapusta et al. (2013)).

TEs são sequências repetitivas com potencial de mobilização no genoma que constituem pelo menos 45% do genoma humano (Human; Sequencing (2001)) e são classificadas com base em seu modo de transposição. Os elementos da classe I (retrotransposons) se transpõem através de um intermediário de RNA usando a transcriptase reversa e incluem elementos nucleares intercalados longos e curtos (LINEs e SINEs) e elementos de repetição terminal longa (LTR). Elementos de classe II (transposons de DNA) não precisam de um intermediário para mobilização, chamado mecanismo de cortar e colar (Bourque et al. (2018)). Sabe-se que os TEs impactam a arquitetura e a regulação do genoma humano de diferentes maneiras. Os TEs ativos servem como fontes contínuas de variação genética, promovendo novas inserções que podem interromper as sequências de codificação ou alterar a expressão de genes próximos (Bennett et al. (2004)). TEs também podem desencadear rearranjos cromossômicos por meio de recombinação ectópica criando mutações, comumente encontradas em células cancerosas (Bodega; Orlando (2014); Abascal et al. (2015)). Devido às implicações das inserções de TE, elas são consideradas altamente mutagênicas e estão associadas a várias etapas de desenvolvimento e progressão do câncer (Criscione et al. (2014)). Além disso, os TEs são ricos em sequências codificantes e regulatórias que também são frequentemente utilizadas como caixas de construção para gerar novas funções celulares, um processo denominado domesticação (Etchegaray et al. (2021); Abascal et al. (2015); Mager et al. (2003)).

Recentemente, foi proposta a hipótese de que os TEs podem atuar como domínios funcionais do lncRNA. Nesta hipótese, fragmentos de TE são cooptados para formar os chamados RIDLs (*Repeat Insertion Domains of LncRNAs*) servindo como sítios de ligação para proteínas ou outros ácidos nucleicos que são essenciais para a função de lncRNA (Johnson; Guigó (2014)). Carlevaro-Fita et al. (2019) forneceu um catálogo de todo o transcriptoma humano de RIDLs putativos identificando 5374 inserções repetitivas em 3566 genes lncRNA que poderiam ser considerados RIDL funcionais entre mais de 40.000 outras inserções exônicas de

passageiros. Existem evidências, validadas em estudos de associação com doenças de que lncRNAs contendo sequências RIDLs possam exercer importantes funções biológicas (Carlevaro-Fita et al., 2020).

Devido à importância dos TEs inseridos em lncRNAs e o impacto dessas sequências repetitivas em células cancerosas, pesquisas por genes de lncRNA-RIDLs expressos diferencialmente em diferentes tipos de câncer podem ajudar a identificar lncRNAs clinicamente relevantes como marcadores moleculares potenciais. Aqui, investigamos genes lncRNA-RIDLs diferencialmente expressos em oito tipos de câncer em comparação com sua contraparte não tumoral usando dados do *The Cancer Genome Atlas* e avaliamos sua utilidade como marcadores de prognóstico em análises de sobrevivência, destacando a importância dessas moléculas em tumores.

2 REVISÃO DE LITERATURA

2.1 Câncer

O câncer é um grande grupo de mais de 100 doenças em que as células anormais crescem descontroladamente, indo além de seus limites habituais para invadir partes adjacentes do corpo e se espalhar para outros órgãos (INCA (2020)). Fatores genéticos e ambientais, incluindo o estilo de vida da população, estão envolvidos na etiologia do câncer, como a exposição a produtos químicos da fumaça do tabaco, ou radiação, como os raios ultravioleta do sol e a dieta (Wogan et al. (2004)).

Conforme a Sung et al. (2021), o câncer é a primeira ou a segunda causa de morte prematura em 134 dos 183 países no mundo, com mais de 18,1 milhões de novos casos e 9,6 milhões de mortes por ano.

De acordo com o INCA (2020), os cânceres que tiveram maior mortalidade no Brasil em 2019 entre os homens foram traqueia, brônquios e pulmões (13,8%), seguido do de próstata (13,1%) e cólon e reto (8,4%); entre mulheres foram o de mama (16,4%), seguido do de traqueia, brônquios e pulmões (11,4%) e cólon e reto (9,4%). Estima-se que em 2020, no Brasil, o câncer mais frequente entre os homens seja o de próstata, seguido do de cólon e reto e traqueia e brônquios e pulmões; já entre as mulheres seja o de mama, seguido do de cólon e reto e colo do útero (INCA (2020)). No mundo, os cânceres com maior mortalidade em 2020 entre os homens foram o de pulmão (21,5%), seguido do fígado (10,5%) e cólon e reto (9,3%); entre mulheres foram o de mama (15,5%), seguido de pulmão (13,7%) e cólon e reto (9,5%) (Sung et al. (2021)).

Em 2020, o câncer de esôfago foi o sexto mais frequente entre os homens no Brasil e o sétimo mais frequente no mundo. O câncer de bexiga é uma das neoplasias mais comuns do trato urinário e o décimo tipo mais incidente, em nível mundial (INCA (2020); Sung et al. (2021). Segundo o CTCA (2020)), o principal tipo de câncer que pode atingir o rim, e que corresponde a 80% dos casos, é o carcinoma renal de células claras. Nele, as células cancerosas normalmente se desenvolvem no revestimento de tubos muito pequenos nos rins, chamados túbulos, e com o tempo, essas células

podem se transformar em uma massa e causar uma obstrução. O prognóstico desse tipo de câncer depende, dentre outros fatores, da idade do paciente e da rapidez em se diagnosticar e tratar a doença.

No século 21, espera-se que o câncer seja classificado como a principal causa de morte e a barreira mais crítica para o aumento da expectativa de vida no mundo (Bray et al. (2018)). A maioria dos tipos de tumores malignos são sintomáticos apenas em um estágio avançado. Desse modo, no momento em que são detectados, já são compostos por bilhões de células com características anormais (Hanahan; Weinberg (2011)).

Recentemente, as taxas de sobrevivência de pacientes na maioria dos tipos de tumor aumentaram, em parte devido à implementação de marcadores moleculares de diagnóstico e prognóstico, que levam a uma melhor assistência ao paciente. Embora a maioria dos marcadores moleculares seja baseada em regiões codificantes do genoma, um número crescente de estudos mostra promissor o aprendizado de regiões não codificantes, como RNAs não codificantes (ncRNAs) (Hon et al. (2019)).

2.2 Longos RNAs não codificantes

Os ncRNAs são geralmente divididos em duas categorias de acordo com seu tamanho: pequenos ncRNAs, transcritos que apresentam menos de 200 nucleotídeos (nt) de comprimento e longos ncRNAs (lncRNAs) maiores que 200 nt. Os pequenos ncRNAs incluem os bem caracterizados microRNAs (miRNAs), que atuam como importantes reguladores pós-transcricionais da expressão gênica via o complexo silenciador induzido por RNA (Kwok; Tay (2017)). Também entre os pequenos ncRNAs, podemos citar: RNA ribossômico (rRNA), RNA transportador (tRNA), pequeno RNA nuclear (snRNA), pequeno RNA nucleolar (snoRNA) e RNA de interação com PIWI (piRNA) (Esteller et al. (2018)). Os lncRNAs são mais frequentemente transcritos pela RNA polimerase II, podem ter íntrons e variantes de *splicing*, podem exibir cauda poli A na extremidade 3' e podem sofrer modificações epigenéticas semelhantes a genes que codificam proteínas, como regulação pela metilação do DNA e mudanças nas histonas (Yao et al. (2019)).

Os lncRNAs ainda são pouco compreendidos e parecem estar envolvidos em diversos processos biológicos, sendo associados com os processos de carcinogênese, progressão tumoral, metástase, proliferação celular, evasão da apoptose e o desenvolvimento de resistência a quimioterápicos (FIGURA 1) (Bhan et al. (2017)). Atualmente já se sabe que alguns lncRNAs também podem codificar pequenos peptídeos e pesquisas funcionais desses peptídeos são de suma importância para a biologia e a medicina (Li et al. (2017)).

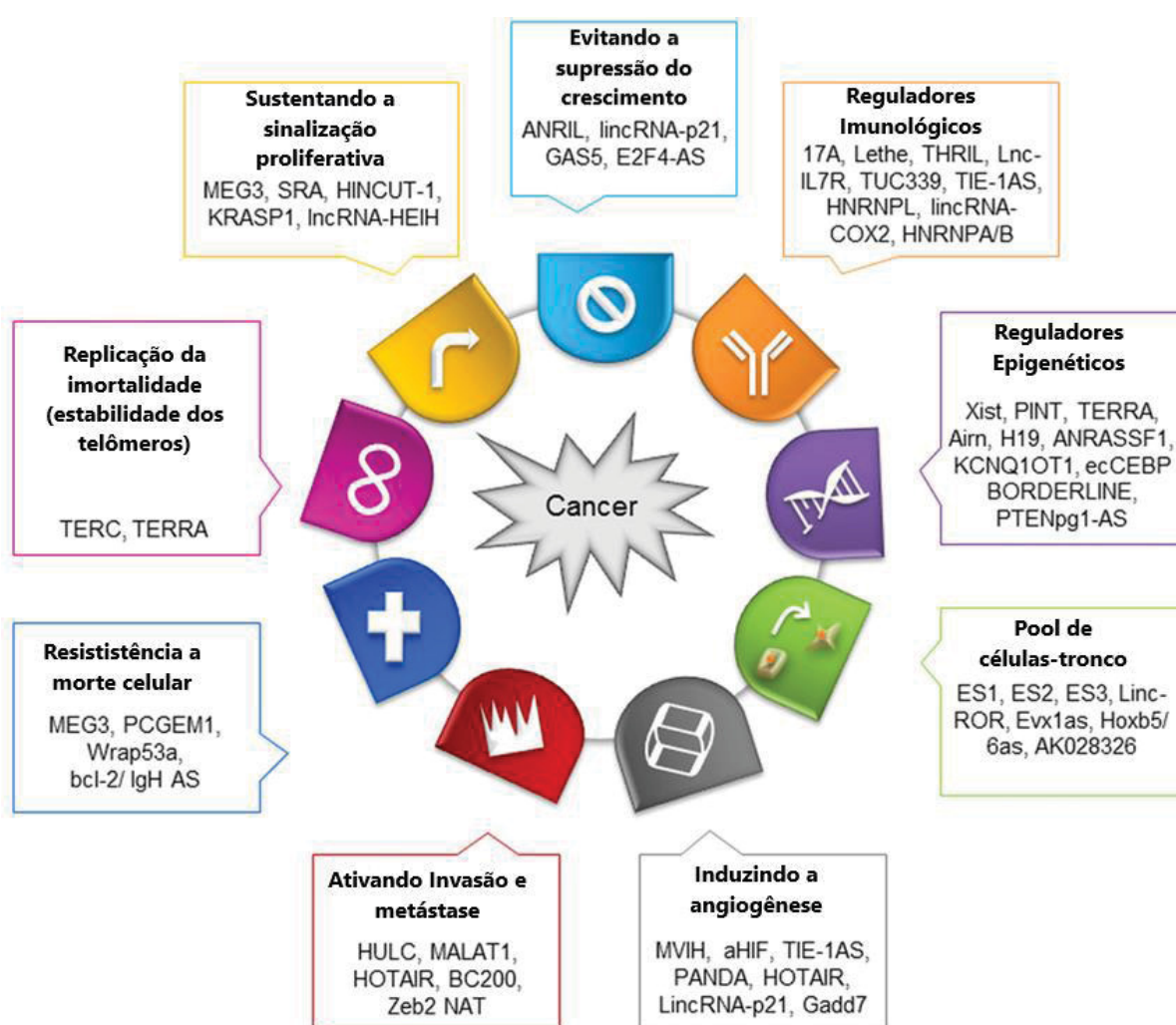


FIGURA 1: LncRNAs associados à carcinogênese. Fonte: adaptado de Parasramka et al. (2016).

Os lncRNAs podem ser classificados em relação a sua posição com a de genes codificantes de proteínas próximos. A partir desta classificação, os lncRNAs podem

ser divididos em cinco grupos principais: intrônicos, intergênicos (lincRNAs), sobreposição-antisense, sobreposição-sense e bidirecional (FIGURA 2) (Wu; Du (2017)).

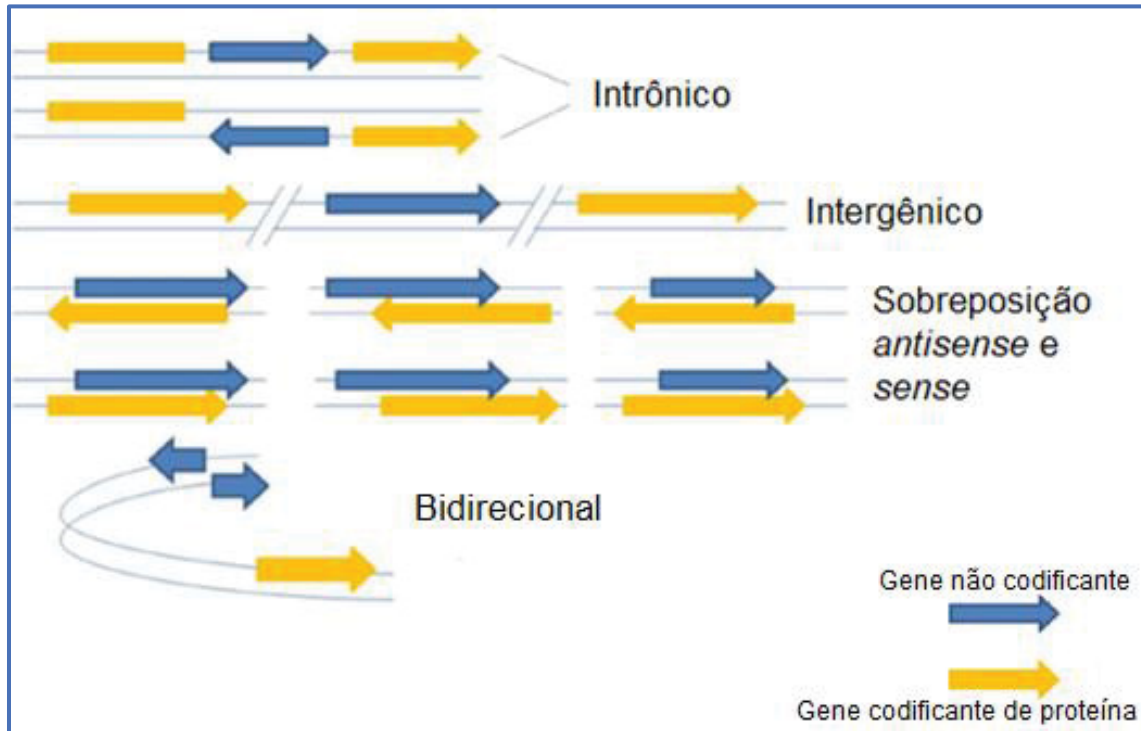
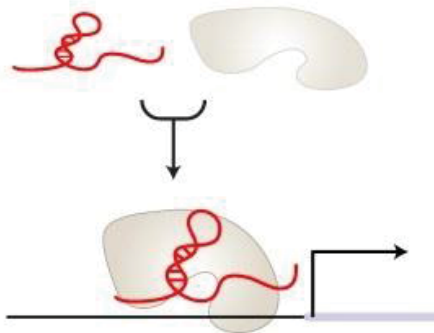


FIGURA 2. Classificação dos lincRNAs em relação a sua posição com a de genes codificantes de proteínas próximos. FONTE: adaptado de Wu; Du (2017).

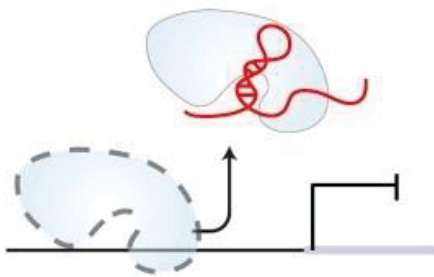
Os lincRNAs podem atuar em cis, interferindo na regulação da expressão de genes próximos ao seu local de síntese, ou atuar em trans, regulando a expressão de genes localizados em outras regiões do genoma (Nguyen et al. (2020) Ferré et al. (2016)). Esta diversidade reflete a própria versatilidade da molécula de RNA: através de seu dobramento em estruturas secundárias, o RNA pode-se ligar a uma grande variedade de substratos de forma altamente específica, podendo interagir com DNA, proteínas e outros RNAs (Geisler; Coller (2013)). Alguns mecanismos de ação dos lincRNAs já foram descritos, eles podem agir como: sinalizador, isca, guia, arcabouço, acentuador de RNAs (eRNAs), remodelagem de cromatina, interações de cromatina e competidor endógeno (ceRNA), podendo atuar em nível transcricional, pós-transcricional e pós-traducional (FIGURA 3) (Fang; Fullwood (2016)).

Epigenético e Transcricional

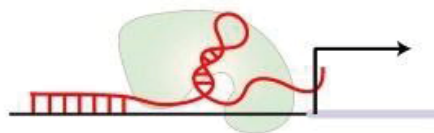
a Ativador Transcricional



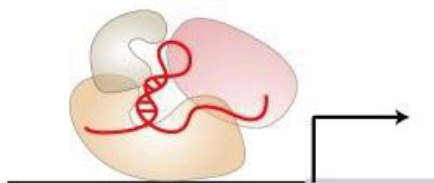
b Repressor Transcricional



c Guia Transcricional

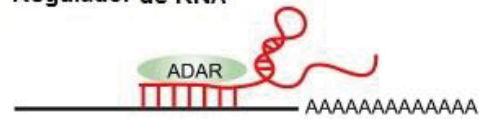


d Arcabouço para Complexos Modificadores de Cromatina

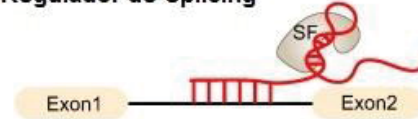


Pós-Transcricional

e Regulador de RNA



f Regulador de Splicing



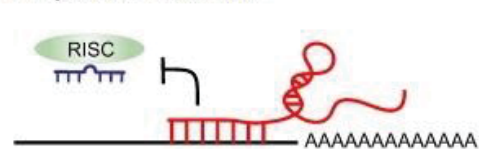
g Ancorador de miRNA



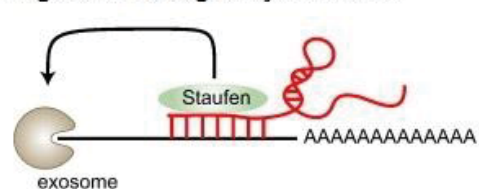
h Sequestrador de miRNA



i Bloqueador de miRNA



j Regulador na degradação do RNA



k Regulador na Eficiência de Tradução

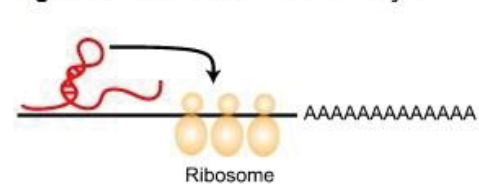


FIGURA 3. Mecanismos de ação de lncRNAs. FONTE: adaptado de Yang et al. (2014).

Já são conhecidos alguns papéis de lncRNAs no câncer. O lncRNA HOTAIR, que é um dos mais famosos, desempenha um papel importante nos cânceres de pulmão, renal, esôfago, ovário, mama e outros tipos (Oliveira et al. (2019)). O lncRNA-MALAT1 está envolvido com o processo de metástase de câncer de pulmão

e pode servir como alvo de tratamento (Liu et al. (2018)) e também já foi observado superexpresso em outros tipos de câncer, como o câncer de mama (Zeng et al. (2018)) e câncer cervical (Zhang et al. (2015)). O lncRNA SNHG15 foi relacionado com diversos cânceres incluindo câncer gástrico, hepatocelular, pancreático, colorretal, mama, tireoide, glioma, osteossarcoma, entre outros. Seu papel é crítico na manutenção da regulação da proliferação celular, migração e invasão de tumores. A sua superexpressão está associada a menor sobrevida (Tong et al. (2019)).

A diferenciação de genes lncRNA reais de ruído transcricional e a caracterização de seus domínios funcionais são úteis para a identificação mais rápida de lncRNAs com papel em processos fisiológicos e doenças. Pouco se sabe sobre a origem desses genes; no entanto, elementos transponíveis (TEs) têm sido associados ao surgimento e diversificação de lncRNA (Hadjigryrou; Delihias (2013); Kapusta et al. (2013)). Conhecer esses domínios de uma maneira abrangente é, portanto, um passo fundamental para a compreensão e previsão da função do lncRNA.

2.3 Elementos transponíveis

Os TEs também conhecidos como transposons ou elementos móveis, são sequências de DNA com potencial de se moverem ou se duplicarem de um local do genoma para outro. Eles constituem a maior parte das sequências repetitivas na maioria dos genomas eucarióticos, representando pelo menos 45% do genoma humano (Human; Sequencing (2001)). Outras sequências repetitivas incluem repetições em tandem (sequências de satélite ou mini-satélites), duplicações genômicas esporádicas e alguns genes hospedeiros de várias cópias (como rRNA, tRNA, genes de histonas etc) (Bao et al. (2015)).

Eles foram descobertos em 1940 por Barbara MacClintock, observando a alteração do padrão de coloração dos grãos de milho após cruzamentos. Deste trabalho foi concluído que existiam “elementos controladores” que possuíam a habilidade de se mover de um lugar para outro do cromossomo (Biémont; Vieira (2006)).

Os elementos de transposição em eucariotos podem ser classificados de duas maneiras. A primeira é quanto a autonomia ou não para realização da transposição. Nesta eles são chamados de autônomo ou não-autônomos, sendo que os primeiros possuem os genes necessários para realização da transposição, e os segundos não os possuem, dependendo diretamente dos primeiros. Os DNAs móveis autônomos codificam proteínas necessárias para seu movimento no genoma, incluindo aquelas com atividade enzimática da transcriptase reversa (classe I) ou transposase (classe II) (Finnegan (1989)).

A segunda maneira é de acordo com seu mecanismo de transposição, ou seja, qual intermediário é utilizado, segundo o sistema de Wicker 2007: elementos de classe I (retrotransposons) transpõem através de um intermediário de RNA usando a transcriptase reversa; eles incluem elementos intercalados longos (LINE), elementos nucleares intercalados curtos (SINEs), retrotransposons semelhantes a Penelope (elementos semelhantes a Penelope - PLE) e elementos repetidos terminais longos (LTR), como retrotransposons DIRS (ou retrotransposons que codificam tirosina recombinase, Retrotransposições YR). Por sua vez, os elementos de classe II (transposons de DNA) se transpõem por meio de um mecanismo de “cortar e colar”, sem uma molécula intermediária (FIGURA 4) (Bourque et al. (2018)).

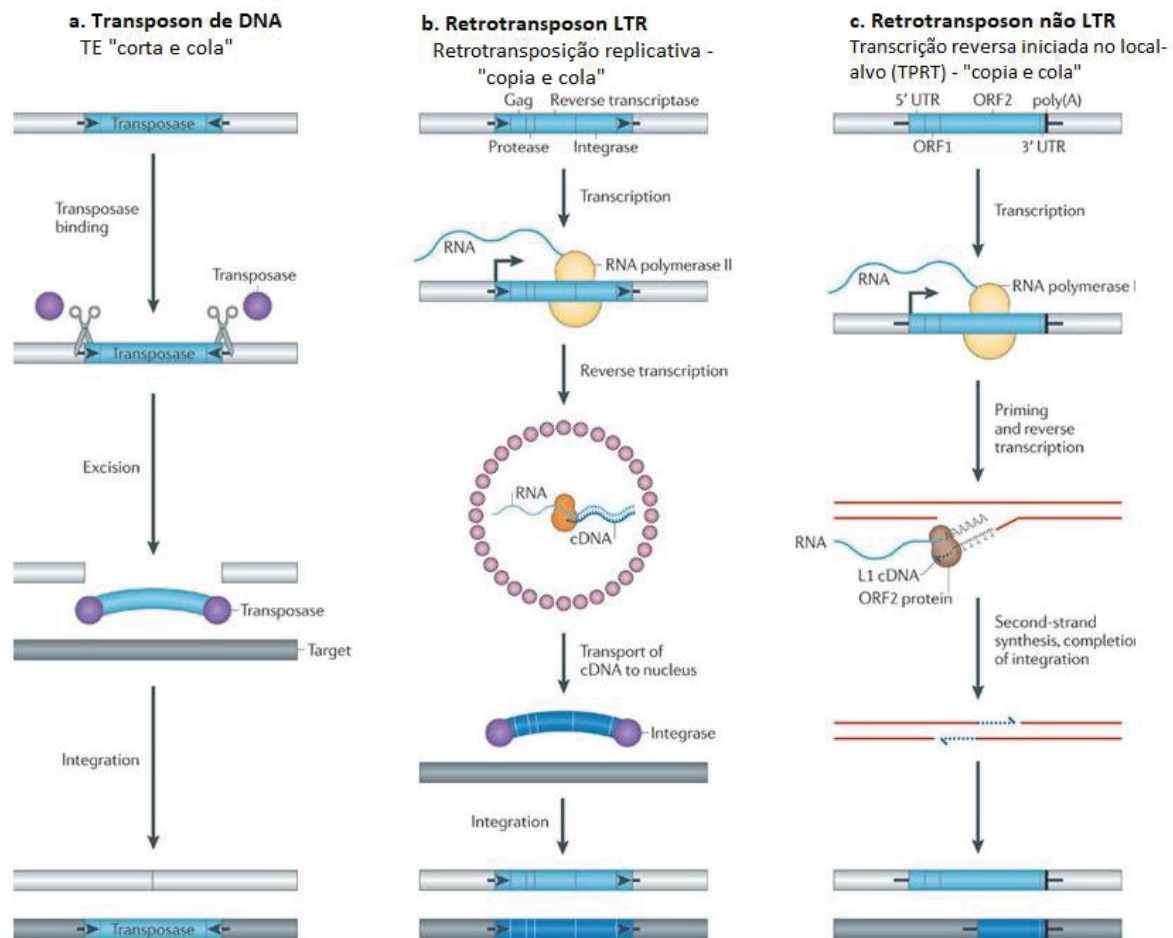


FIGURA 4. Os diversos mecanismos de mobilização do transposon. **a.** Transposons de DNA. Muitos transposons de DNA são flanqueados por repetições terminais invertidas (TIRs; setas pretas), codificam uma transposase (círculos roxos) e se mobilizam por um mecanismo de 'cortar e colar' (representado pela tesoura). A transposase se liga nos TIRs ou perto deles, retira o transposon de sua localização genômica existente (barra cinza claro) e cola-o em uma nova localização genômica (barra cinza escuro). As clivagens das duas fitas no local alvo são escalonadas, resultando em uma duplicação do local alvo (TSD) tipicamente de 4-8 pares de base (pb) (linhas pretas horizontais curtas flanqueando o elemento transponível (TE)). **b.** Os retrotransposons LTR contêm duas repetições terminais longas (LTRs; setas pretas) e codificam as atividades Gag, protease, transcriptase reversa e integrase, todas cruciais para a retrotransposição. O 5' LTR contém um promotor que é reconhecido pela RNA polimerase II do hospedeiro e produz o mRNA do TE (o local de início da transcrição é indicado pela seta em ângulo reto). Na primeira etapa da reação, as proteínas Gag (pequenos círculos rosa) se agrupam em partículas semelhantes a vírus que contêm mRNA de TE (azul claro), transcriptase reversa (formato laranja) e integrase. A transcriptase reversa copia o mRNA de TE em um dsDNA de comprimento total. Na segunda etapa, a integrase (círculos roxos) insere o cDNA (mostrado pelo arco largo e azul escuro) no novo local de destino. Da mesma forma que as transposases de transposons de DNA, as integrases de retrotransposons criam cortes escalonados nos locais alvo, resultando em TSDs. **c.** Retrotransposons não LTR não possuem LTRs e codificam um ou dois ORFs. Nesse

mecanismo gera um mRNA de comprimento total (ondulado, linha azul claro). No entanto, esses elementos se mobilizam por transcrição reversa iniciada no local-alvo (TPRT). Nesse mecanismo, uma endonuclease codificada por elemento gera uma sequência de fita simples no DNA genômico, liberando um 3'-OH que é usado para iniciar a transcrição reversa do RNA. As proteínas que são codificadas por retrotransposons não LTR autônomos também podem mobilizar RNAs retrotransposons não autônomos, bem como outros RNAs celulares. O mecanismo TPRT de um elemento intercalado longo 1 (L1) está representado na figura; o novo elemento (retângulo azul escuro) é 5' truncado e está com defeito de retrotransposição. Alguns retrotransposons não LTR não possuem caudas poli (A) em suas extremidades 3'. A integração de retrotransposons não LTR pode levar a TSDs ou pequenas deleções no local alvo no DNA genômico. Por exemplo, L1s são geralmente flanqueados por TSDs de 7–20 pb. FONTE: adaptado de Levin et al. (2011).

Kapitonov; Jurka (2007), classificam os TEs por um sistema hierárquico para criar um modelo para implementar a classificação do Wicker dentro do rebase (banco de dados de repetitivos), que reconcilia a aplicação de mecânica e critérios enzimáticos, subdividindo em: classe, subclasse, ordem, superfamília, família e subfamília. Exemplo: o TE Alu é da classe I (retrotransposon), subclasse retrotransposon sem LTR, ordem SINE, superfamília SINE1/7SL, família Alu, subfamília AluSq2, de origem primata.

Sabe-se que os TEs impactam a arquitetura e a regulação do genoma humano de diferentes maneiras. Os TEs ativos servem como fontes contínuas de variação genética, promovendo novas inserções que podem interromper as sequências de codificação ou alterar a expressão de genes próximos (Bennett et al. (2004)). TEs também podem desencadear rearranjos cromossômicos por meio de recombinação ectópica criando mutações, comumente encontradas em células cancerosas (Abascal et al. (2015); Bodega; Orlando (2014)). Devido às implicações das inserções de TE, elas são consideradas altamente mutagênicas e estão associadas a várias etapas de desenvolvimento e progressão do câncer (Criscione et al. (2014)). Além disso, os TEs são ricos em sequências codificantes e regulatórias que também são frequentemente utilizadas como caixas de construção para gerar novas funções celulares, um processo denominado domesticação (Etchegaray et al. (2021); Bourque et al. (2018); Abascal et al. (2015); Mager et al. (2003)).

2.4 Elementos RIDLs

Recentemente, foi proposta a hipótese de que os TEs podem atuar como domínios funcionais do lncRNA. Nesta hipótese, fragmentos de TE são cooptados para formar os chamados RIDLs (*Repeat Insertion Domains of LncRNAs*) servindo como sítios de ligação para proteínas ou outros ácidos nucleicos que são essenciais para a função de lncRNA (Johnson; Guigó (2014)). Carlevaro-Fita et al. (2019) forneceu um catálogo de todo o transcriptoma humano de RIDLs putativos identificando 5.374 inserções repetitivas em 3.566 genes lncRNA que poderiam ser considerados RIDL funcionais entre mais de 40.000 outras inserções exônicas. A inserção de TE é um mecanismo evolutivo geral pelo qual a funcionalidade do lncRNA evolui por meio da adição combinatória de domínios distintos de TE que resultam em propriedades emergentes e complexas em seu lncRNA hospedeiro (FIGURA 5) (Johnson; Guigó (2014)).

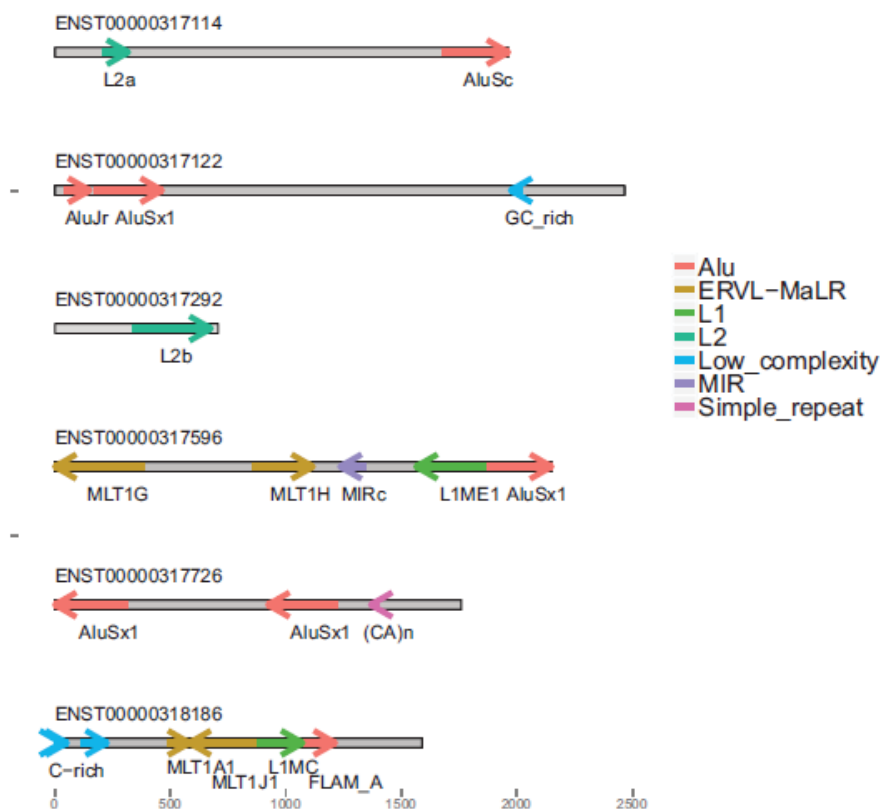


FIGURA 5. Exemplos de perfis de inserção de DNAs repetitivos em lncRNAs. As inserções são representadas por setas, coloridas pela classe TE. Os retângulos representam transcritos de lncRNAs maduros e estão em escala. Fonte: adaptado de Johnson; Guigó (2014).

A localização subcelular e os domínios que a controlam são determinantes cruciais das funções lncRNAs, onde os domínios funcionais de lncRNAs podem recrutar complexos modificadores da cromatina ou fatores de transcrição, ou outra conformação estrutural de lncRNAs que pode, subsequentemente, impedir que essas moléculas de lncRNAs sejam exportadas para o citoplasma e facilitar sua localização nuclear em cis ou em trans (Chen (2016)). Com isso, qualquer RIDL que opere por meio de pareamento de base com a sequência de nucleotídeos complementar, seja DNA ou RNA, ou moléculas de proteína, terá uma infinidade de locais de ligação potenciais em todo o genoma (FIGURA 6) (Carlevaro-Fita et al. (2019); Johnson; Guigó (2014)).

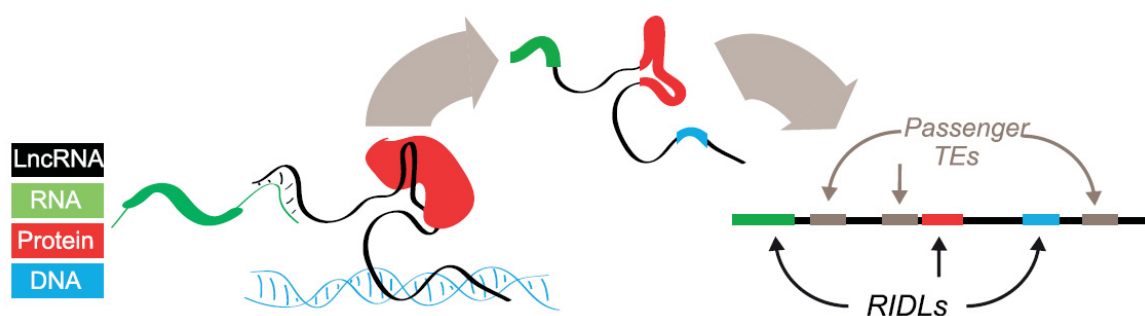


FIGURA 6. Modelo RIDL. Os fragmentos de TEs inseridos exonicamente contêm atividades pré-formadas de ligação a proteínas (vermelho), ligação a RNA (verde) ou ligação a DNA (azul) que contribuem para a funcionalidade do lncRNA do hospedeiro (preto). É provável que os RIDLs sejam uma pequena minoria de TEs exônicos, coexistindo com um grande número de “passageiros” não funcionais (cinza). FONTE: adaptado de Carlevaro-Fita et al. (2019).

Um exemplo de lncRNA com RIDL que tem função de oncogene é o *Plasmacytoma variant translocation 1* (PVT1). Segundo Carlevaro-fita et al. (2019), este lncRNA contém um RIDL (MIR – SINE). Ele interfere com a fosforilação da mielocitomatose (MYC) em Thr58 no núcleo, o que subsequentemente aumenta a estabilidade de MYC, que tem efeitos oncogênicos em vários cânceres (Li et al. (2019)).

Devido à importância dos TEs em formar lncRNAs funcionais e o impacto dessas sequências repetitivas em células cancerosas, pesquisas por genes de lncRNA-

RIDLs expressos diferencialmente no câncer podem ajudar a identificar lncRNA-RIDLs clinicamente relevantes como marcadores moleculares potenciais. Aqui, investigamos genes de lncRNA-RIDLs expressos diferencialmente em oito tipos de câncer (carcinoma urotelial de bexiga - BLCA; carcinoma de mama invasivo - BRCA; adenocarcinoma de cólon - COAD; carcinoma de esôfago - ESCA; carcinoma renal de células claras - KIRC; carcinoma hepático - LIHC; adenocarcinoma pulmonar - LUAD; e adenocarcinoma de próstata – PRAD) em comparação com sua contraparte não tumoral usando dados do *The Cancer Genome Atlas* e avaliamos sua utilidade como marcadores de prognóstico em análises de sobrevida, destacando a importância dessas moléculas em tumores.

3 OBJETIVOS

3.1. OBJETIVO GERAL

Avaliar a diferença de expressão e associação com prognóstico de genes lncRNA-RIDLs no câncer.

3.2. OBJETIVOS ESPECÍFICOS

- Identificar lncRNAs diferencialmente expressos em BLCA, BRCA, COAD, ESCA, KIRC, LIHC, LUAD e PRAD em relação a sua contraparte não tumoral por meio de dados depositados no banco de dados TCGA;
- Identificar, entre os lncRNAs diferencialmente expressos em câncer, quais são lncRNA-RIDLs e evidenciar os domínios TEs que apresentam;
- Investigar a associação entre os genes lncRNA-RIDLs diferencialmente expressos e a sobrevida dos pacientes com BLCA, BRCA, COAD, ESCA, KIRC, LIHC, LUAD e PRAD; e
- Avaliar a frequência de lncRNA-RIDLs diferencialmente expressos entre os lncRNAs descritos no *Cancer LncRNA Census* com reconhecida importância no câncer.

4 MATERIAL E MÉTODOS

4.1. Extração de dados e análise de expressão diferencial

Os dados de 14.660 lncRNAs de 8 tipos de câncer (BLCA, BRCA, COAD, ESCA, KIRC, LIHC, LUAD e PRAD) foram baixados do *The Cancer Genome Atlas* (TCGA) usando o pacote R *GDCRNA Tools* (Li et al. (2018)). As análises de expressão diferencial foram realizadas comparando cada tipo de câncer com sua contraparte não tumoral usando *limma pipeline* (Ritchie et al. (2015)). Usamos como critério de filtro o valor de $p < 0,01$ para selecionar os lncRNAs expressos diferencialmente.

4.2. Identificação dos genes lncRNA-RIDLs expressos diferencialmente em cânceres

Para este trabalho, utilizamos os dados RIDL fornecidos por Carlevaro-Fcita et al. (2019). A tabela de anotação RIDL fornecida por esses autores foi primeiro filtrada para eliminar as sequências não-TE, excluindo as repetições classificadas como “baixa complexidade”, “satélite” e “repetição simples”, restando 4.556 repetições. A anotação do lncRNA foi baixada do GENCODE v21 - www.gencodegenes.org (Harrow et al. (2012)), a mesma versão usada pelos autores anteriores. Para obter as informações de todos os genes lncRNA-RIDLs, comparamos ambas as tabelas usando um código Python interno escrito para identificar os genes lncRNAs que contêm sequências RIDL com base nas coordenadas cromossômicas de cada tabela.

Todos os lncRNAs diferencialmente expressos encontrados em cada câncer foram interceptados com a tabela de genes lncRNA-RIDLs (Tabela complementar 2). A interseção foi analisada usando a ferramenta Diagramas de Venn (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Nós avaliamos a sobre-representação dos genes lncRNA-RIDLs entre os tipos de câncer usando o teste hipergeométrico. Isso foi realizado com o software R e considerado valor de $p < 0,05$. O mapa de calor foi construído usando o pacote Complex Heatmap R (Gu et al. (2016)).

4.3. Análise de Sobrevida

Os dados de sobrevida de todos os tipos de câncer analisados foram baixados do banco de dados do Xena Browser (<https://xenabrowser.net/datapages/>) e a análise de regressão de riscos proporcionais de Cox univariada foi realizada usando o pacote R “sobrevivência”. Consideramos significativo o valor de p de cox $<0,05$.

4.4. Identificação de genes lncRNA-RIDLs com um papel significativo no câncer

Para identificar genes lncRNA-RIDLs com papel significativo no câncer, usamos a lista de 122 lncRNAs GENCODE com papéis causais em fenótipos de câncer representados no *Cancer LncRNA Census* (CLC) por Carlevaro-fita et al. (2020).

5 CAPÍTULO I

O artigo foi submetido para a revista *Clinical Genetics*.

Article Title: LNCRNAS CONTAINING TRANSPOSABLE ELEMENT SEQUENCES IN CANCER: RIDL AS NEW BIOMARKERS

Article Type: Original Article

Date Submitted: 12-May-2021

Manuscript ID: CGE-00379-2021

Authors:

First author Branca Engel Timoner, Department of Genetics, Universidade Federal do Paraná, Curitiba, PR, Brazil. brancatimoner@gmail.com
Second author Carolina Mathias, Department of Genetics, Universidade Federal do Paraná, Curitiba, PR, Brazil. carol.mathias1@hotmail.com
Third author Erika Pereira Zambalde, Department of Genetics, Universidade Federal do Paraná, Curitiba, PR, Brazil. erikazambaldi@gmail.com
Fourth author Leandro Encarnação Garcia, Department of Genetics, Universidade Federal do Paraná, Curitiba, PR, Brazil. leandrogarcia20504@gmail.com
Fifth author Daniela Fiori Gradia, Department of Genetics, Universidade Federal do Paraná, Curitiba, PR, Brazil. danielagradia@gmail.com
Sixth author Adriana Ludwig, Fundação Oswaldo Cruz-Fiocruz, Instituto Carlos Chagas, Laboratório de Ciências e Tecnologias Aplicadas em Saúde, Curitiba, PR, Brazil. adriludwig@gmail.com
Last author * Jaqueline Carvalho de Oliveira, Department of Genetics, Universidade Federal do Paraná, Curitiba, PR, Brazil. jaqueline.genetica@gmail.com*

Highlights

- transposable elements (TEs) can act as functional lncRNA domains;
- Repeat Insertion Domains of lncRNAs (RIDL) is still underexplored in cancer;
- Evaluated lncRNAs-RIDLs differentially expressed in eight cancer types using data from The Cancer Genome Atlas;
- 944 differentially expressed lncRNA-RIDLs genes in cancer;
- In the top 20 most differentially expressed, 44% were associated with survival time in at least one tumor type;
- among a compilation of 122 lncRNAs with causal roles in cancer phenotype, 15.6% had RIDLs;
- highlight lncRNA-RIDLs genes as potential biomarkers in cancer.

Abstract

Cancer cells often accumulate multiple mutations and gene expression abnormalities, including in long non-coding RNAs (lncRNAs). Most lncRNAs, however, are still underexplored. It has been recently proposed that transposable elements (TEs) can act as functional lncRNA domains. In this hypothesis, TE fragments are co-opted to form RIDLs (Repeat Insertion Domains of lncRNAs). We investigated lncRNAs differentially expressed in eight types of cancer compared to their non-tumoral counterparts, using data from The Cancer Genome Atlas. Among them, we selected lncRNAs containing sequences of RIDLs (lncRNA-RIDLs). We found 944 differentially expressed lncRNA-RIDLs genes in cancer. Three lncRNA-RIDLs stood out in at least seven types of cancer: SNHG3, MAGI2-AS3, and PVT1. In the top 20 most differentially expressed lncRNA-RIDLs genes, 44% were associated with survival time in at least one tumor type; the ones related to survival in most cancer types were CARMN, LINC00261, LINC00511, and MIR99AHG. Additionally, among a compilation of 122 lncRNAs with causal roles in cancer phenotype, 19 (15.6%) had RIDLs. Our study contributes to the investigation of lncRNA-RIDLs genes and indicates new possibilities for the identification of biomarkers in cancer.

Keywords: lncRNAs, RIDL, transposable elements, differential expression, molecular marker, cancer.

1. INTRODUCTION

Cancer is either the first or second leading cause of premature death in 134 of 183 countries, with more than 18.1 million new cases and 9.6 million deaths per year¹. The development of this disease involves the acquisition of certain traits by cancer cells, such as resistance to growth suppression, evasion of the immune system, sustained proliferative signaling, replicative immortality, promotion of angiogenesis, invasion and metastasis, dysregulation of cellular metabolism, induction of inflammation, genomic instability, and epigenetic changes².

Recently, survival rates of patients with most cancer types have improved, in part due to the identification of diagnostic and prognostic molecular markers. Although most molecular markers are based on coding genomic regions, an increasing number of studies show the suitability of those located in non-coding regions, such as non-coding RNAs^{3,4}.

Rapid technological advances in high-throughput sequencing have revealed thousands of new classes of non-coding RNAs, including the long non-coding RNAs (lncRNAs). lncRNAs are transcripts longer than 200 nucleotides which are most often transcribed by RNA polymerase II; they can contain introns, display splicing variants, exhibit poly(A) tail at their 3' ends, and undergo epigenetic changes similar to protein-coding genes, such as regulation by DNA methylation and histone modifications⁵. They may be located in both the nucleus and the cytoplasm and participate in gene expression regulation at different levels (transcriptional, post-transcriptional and post-translational)^{6,7}.

Tens of thousands of lncRNA genes have been cataloged⁸ so far, but most are still poorly understood. Distinguishing real lncRNA genes from transcriptional noise and characterizing their functional domains would facilitate the fast identification of lncRNAs with a role in physiological processes and diseases. Little is known about the origins of these genes; however, transposable elements (TEs) have been associated with the emergence and diversification of lncRNAs^{9,10}.

TEs are mobile repetitive sequences that make up at least 45% of the human genome¹¹. They are classified into two main groups based upon their transposition mechanism. Class I elements (retrotransposons) transpose through an RNA intermediate using reverse transcriptase; they include long interspersed elements (LINE), short interspersed nuclear elements (SINEs), Penelope-like retrotransposons (Penelope-like elements - PLE), and long terminal repeat elements (LTR), such as DIRS retrotransposons (or tyrosine recombinase-encoding retrotransposons, YR retrotransposons). In turn, class II elements (DNA transposons) transpose through a "cut-and-paste" mechanism, without an intermediate molecule¹². It is known that TEs impact human genome architecture and gene regulation in different ways. Active TEs serve as a source of genetic variation, promoting novel insertions that could disrupt coding sequences or alter the expression of nearby genes¹³. TEs can also trigger chromosome rearrangements through ectopic recombination, creating mutations, commonly found in cancer cells^{14,15}. Due to the impacts of TE insertions, they are considered

highly mutagenic and are associated with multiple steps of cancer development and progression¹⁶. Moreover, TEs are rich in coding and regulatory sequences that are also frequently used as building blocks to develop new cellular functions, in a process called domestication^{17,15,12,18}.

It has been recently proposed that TEs can act as functional lncRNA domains. According to this hypothesis, TE fragments are co-opted to form Repeat Insertion Domains of lncRNAs (RIDLs), which serve as binding sites for proteins or other nucleic acids that are essential for lncRNA function¹⁹. Carlevaro-Fita et al. provided a human transcriptome-wide catalog of putative RIDLs, identifying 5374 repetitive insertions in 3566 lncRNA genes that could be considered functional RIDLs, among more than 40,000 other exonic “passenger” insertions. Interestingly, the lncRNAs containing RIDLs sequences (lncRNA-RIDLs) are usually functionally validated and associated with diseases²⁰.

Due to the importance of TEs in functional lncRNAs and the impact of these repetitive sequences in cancer cells, the discovery of lncRNA-RIDLs genes differentially expressed in various cancer types can help to identify clinically relevant lncRNAs as potential molecular markers. Herein, we investigated lncRNA-RIDLs differentially expressed in eight types of cancer compared to their non-tumoral counterparts, using data from The Cancer Genome Atlas, and evaluated their utility as prognostic markers in survival analysis.

2. MATERIALS AND METHODS

Data extraction and differential expression analysis

Data consisting of 14,660 lncRNAs from 8 cancer types (bladder urothelial carcinoma – BLCA; invasive breast carcinoma – BRCA; colon adenocarcinoma – COAD; esophageal carcinoma – ESCA; clear-cell renal cell carcinoma – KIRC; liver hepatocellular carcinoma – LIHC; lung adenocarcinoma – LUAD; and prostate adenocarcinoma - PRAD) were downloaded from The Cancer Genome Atlas (TCGA) using the R package GDCRNATools²¹. Differential expression analyses were performed comparing each cancer type to its non-tumoral

counterpart using the limma pipeline²². We used a p-value of <0.01 as our filtering criterion to select differentially expressed lncRNAs.

Identification of differentially expressed lncRNA-RIDL genes in cancers

For this study, we used the RIDL annotation provided by Carlevaro-Fita et al.²⁰. The data was first filtered to remove non-TE sequences, excluding those classified as “low-complexity”, “satellite”, and “simple-repeat”; 4556 repeats remained. We employed the same version of lncRNA annotation used by these authors and downloaded it from GENCODE v21 (www.gencodegenes.org²³). To obtain the data of all the lncRNA-RIDLs genes, we compared both tables using an in-house Python script, developed to identify the lncRNA genes that contain RIDL sequences based on the chromosome coordinates from each table.

All differentially expressed lncRNAs found in each cancer type were intersected with the lncRNA-RIDL genes table (Supplementary Table 2). The intersection was analyzed using Venn Diagrams (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). We evaluated the over-representation of lncRNA-RIDLs genes among cancer types using the hypergeometric test, which was performed using the R software (p-value <0.05). The heatmap was constructed using the Complex Heatmap R package²⁴.

Survival Analysis

Survival data of all analyzed cancer types was downloaded from the Xena Browser database (<https://xenabrowser.net/datapages/>). Univariate Cox proportional hazards regression analysis was performed using the R package “survival”. A cox p-value < 0.05 was considered statistically significant.

Identification of lncRNA-RIDL genes with a significant role in cancer

To identify lncRNA-RIDL genes with a significant role in cancer, we used a list of 122 GENCODE lncRNAs with causal roles in cancer phenotypes compiled in the Cancer lncRNA Census (CLC) by Carlevaro-Fita et al.²⁵.

3. RESULTS

lncRNAs differential expression analysis

To investigate the expression of lncRNA-RIDLs genes in cancer, we first investigated lncRNAs that are differentially expressed among eight cancer types (BLCA, BRCA, COAD, ESCA, KIRC, LIHC, LUAD AND PRAD) using expression data from 14,660 lncRNAs from the Cancer Genome Atlas (TCGA) data. In this analysis, we found a total of 3,509 distinct lncRNAs differentially expressed between tumor and non-tumor counterpart, some of them were modulated in multiple tumor types (Figure 1A). Considering each cancer type, we found 1,671 lncRNAs differentially expressed in BLCA; 375 in BRCA; 388 in COAD; 1,283 in ESCA; 573 in kidney KIRC; 269 in LIHC; 408 in LUAD; and 423 in PRAD (Table 1). Detailed information about differentially expressed lncRNAs in the analyzed cancer types is summarized in Supplementary Table 1.

lncRNA-RIDLs genes in cancers

Carlevaro-Fita et al. presented a catalog of putative RIDLs¹⁹. To distinguish RIDLs from ordinary TE fragments, the authors used three types of selection signature: exonic enrichment, strand bias (based on the host gene), and exon-specific evolutionary conservation. The RIDLs represent a small fraction among many phenotypically neutral “passenger” TEs in lncRNA exons and highlight TEs under evolutionary selection for functions executed by the mature RNA product. The authors started the analysis from 26,414 lncRNA transcripts and reached a final list of 3,566 lncRNA genes that contain 5,375 RIDLs. From

this RIDL list, we removed non-TE repeats (satellites, low-complexity, and simple repeats), resulting in 3,330 lncRNA genes containing 5,023 RIDLs derived from TEs (Figure 1; Supplementary Table 2).

Then, we intersected all those differentially expressed lncRNAs found in each cancer type with these 3,330 lncRNA genes containing TE RIDLs obtaining 944 lncRNA-RIDLs differentially expressed in cancer (Figure 1; and Table 1) highlighting that 26.90% of lncRNAs differentially expressed in cancer contain TE repeats in their exons that are putative functional domains. The list of all lncRNA-RIDLs in each cancer type is available in Supplementary Table 3. A hypergeometric test indicates the overlap between the two lists is not by chance, emphasizing the relevance of RIDL in lncRNAs differentially expressed in cancer (p-value <0.01).

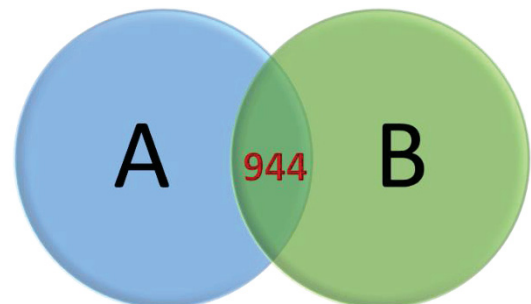
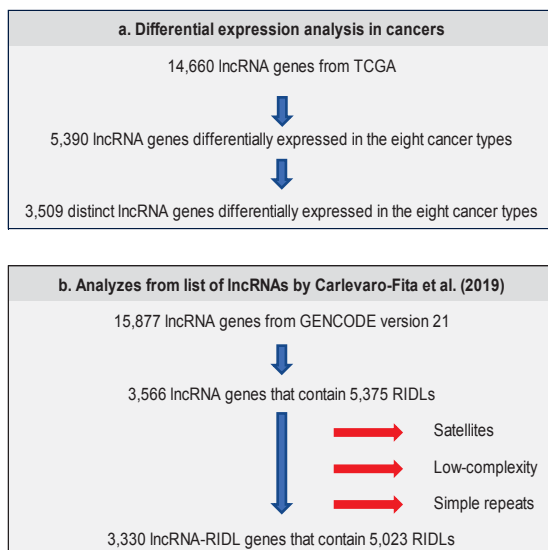


Figure 1. Schematic representation of the identification of lncRNA-RIDLs genes differentially expressed in the eight types of cancer analyzed in this work. A. Summary of analyzed data where “a” represents the differential expression analysis starting from 14,660 lncRNA genes that have their expression analyzed in eight types of cancer, summing 5,390 lncRNA genes differentially expressed of which 3,509 are distinct. In “b” the RIDL analysis of Carlevaro-Fita et al. is represented starting from 15,877 lncRNA genes from which 3,566 contains at least one RIDL (totalizing 5,375 RIDLs)¹⁹. After filtering non-TE repeats, we obtained a dataset of 3,330 lncRNA-RIDL containing at least one RIDL derived from TEs (totalizing 5,023 RIDLs). B. Intersection of lists a and b showing that

944 lncRNAs are common corresponding to lncRNA-RIDLs genes differentially expressed in the eight types of cancer.

An important result to highlight is that the number of exclusive lncRNA-RIDLs genes deregulated to each type of cancer varied considerably, as shown in Table 1. Around 20 exclusive lncRNA-RIDLs genes were found in BRCA, LIHC, LUAD and PRAD, while a higher number was identified in the others, reaching 251 genes in BLCA. In terms of percentage, BRCA and LUAD have the lowest proportion of exclusive lncRNA-RIDLs for their group (12.9% and 13.8%, respectively) whereas BLCA and ESCA - which have the highest number of total lncRNA-RIDLs - have the highest proportion (60.3% and 56.8%, respectively). The existence of these exclusive differentially expressed lncRNA-RIDLs genes highlights the importance to look at specific lncRNAs in these cancers.

Table 1: Number of lncRNA genes differentially expressed in eight cancer types, their respective intersections with the lncRNA-RIDLs, and the number of exclusive lncRNA-RIDLs for each cancer type.

Cancers		Number of lncRNAs	Number of lncRNAs-RIDLs	Number of exclusive lncRNAs-RIDLs (%)
BLCA	Bladder urothelial carcinoma	1671	416	251 (60.3%)
BRCA	Breast invasive carcinoma	375	132	17 (12.9%)
COAD	Colon adenocarcinoma	388	134	30 (22.4%)
ESCA	Esophageal carcinoma	1283	317	180 (56.8%)
KIRC	Kidney renal clear cell carcinoma	573	201	50 (24.9%)
LIHC	Liver hepatocellular carcinoma	269	85	22 (25.9%)
LUAD	Lung adenocarcinoma	408	152	21 (13.8%)
PRAD	Prostate adenocarcinoma	423	119	21 (17.6%)

BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; KIRC: kidney renal clear cell carcinoma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; PRAD: prostate adenocarcinoma.

We found only one lncRNA-RIDL gene that was differentially expressed in all cancer types analyzed - the lncRNA SNHG3. Moreover, we found other two that appear in seven cancer groups: MAGI2-AS3 and PVT1. Although there is

variation in the level of expression (represented by the log2FC values) among tumor groups, each of these lncRNA-RIDLs genes has a concordant direction of expression among tumors compared to its non-tumoral counterpart - SNHG3 and PVT1 are up-regulated while MAGI2-AS3 is down-regulated in all cancer types.

We separate the up- and down-regulated lncRNA-RIDLs genes and then analyze the common ones between cancer types (Table 2).

Table 2: Number of lncRNA-RIDLs genes that were downregulated (below and left in the diagonal) and upregulated (above and right in the diagonal) in the eight cancer types and the pairwise comparisons of common lncRNA-RIDLs genes between types of cancers.

	Total down	BLCA	BRCA	COAD	ESCA	KIRC	LIHC	LUAD	PRAD	Total up
BLCA	241		10	10	38	13	10	19	19	175
BRCA	97	30		4	5	7	5	13	15	35
COAD	66	10	22		4	21	15	22	20	68
ESCA	226	55	20	10		12	4	8	5	91
KIRC	85	16	31	21	19		14	27	31	116
LIHC	34	6	6	7	7	6		17	14	51
LUAD	68	23	28	18	14	15	5		25	84
PRAD	32	14	16	6	10	6	3	8		87

BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; KIRC: kidney renal clear cell carcinoma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; PRAD: prostate adenocarcinoma.

The proportion of down- and upregulated lncRNA-RIDLs genes varied among cancer types. For example, around 70% of deregulated lncRNA-RIDLs of BRCA and ESCA were down-regulated, whereas for PRAD the opposite occurred, around 70% were up-regulated. For KIRC and LIHC around 60% were also upregulated. The remaining cancer types have a less discrepant proportion between expression groups.

Reflecting the total number of deregulated lncRNAs, BLCA and ESCA are the cancer types that share the highest number of up- (38) and down-regulated (55) lncRNA-RIDLs genes, while comparisons among BRCA, COAD, LIHC,

LUAD and PRAD showed the smallest number of shared up- and downregulated genes.

Using the differential expression data of lncRNA-RIDLs genes (Supplementary Table 3), we focused on the 20 genes that are most differentially expressed in each cancer type (Table 3) to investigate whether these lncRNA-RIDLs genes could them.

Table 3: Top 20 differentially expressed lncRNA-RIDLs genes in each analyzed cancer type.

BLCA		BRCA		COAD		ESCA	
lncRNA-RIDL	log2FC	lncRNA-RIDL	log2FC	lncRNA-RIDL	log2FC	lncRNA-RIDL	log2FC
AF165147.1	-5.07	CARMN	-4.24	FENDRR	-2.17	LINC01105	-5.19
AC135012.3	-4.98	MIR99AHG	-2.53	LINC02038	-2.14	LINC00982	-5.16
HAND2-AS1	-4.93	SOX9-AS1	-2.50	LINC00483	-2.01	LINC00940	-4.86
FENDRR	-4.67	EMX2OS	-2.39	SMIM2-AS1	-1.92	LINC00261	-4.13
ADAMTS9-AS2	-4.56	MAGI2-AS3	-2.37	LHFPL3-AS2	-1.76	AP004608.1	-3.73
AF001548.1	-4.44	LINC01140	-2.35	AP002761.4	-1.66	AC079061.1	-3.48
TARID	-4.24	AL031587.6	-2.11	USP30-AS1	-1.64	AC012409.3	-3.42
CARMN	-4.09	FZD10-AS1	-1.98	LINC00654	-1.58	CADM3-AS1	-3.29
CADM3-AS1	-3.97	LINC00987	-1.88	TMEM147-AS1	1.59	TUSC8	-3.19
AL354861.3	-3.92	MIR205HG	-1.83	AC016831.1	1.89	AL358176.4	-3.15
FRMD6-AS2	-3.85	PART1	-1.83	FOXP4-AS1	2.16	AC015908.3	-3.12
AC134312.1	-3.77	KLF3-AS1	-1.79	LINC01979	2.45	AC005920.4	-3.11
AC079313.2	-3.69	AC015712.2	-1.74	CASC8	2.48	AC103702.2	3.09
MEF2C-AS1	-3.63	PWAR6	-1.63	AL391056.1	2.50	DUXAP8	3.10
AP003071.4	-3.57	LINC00641	-1.60	PVT1	2.58	LINC01633	3.14
LINC01081	-3.55	AL021392.1	1.60	UCA1	2.94	LINC00460	3.26
AP000911.1	-3.43	LINC01977	1.70	LINC01978	3.01	AL513123.1	3.35
DIO3OS	-3.34	LINC00511	2.06	ELFN1-AS1	4.54	AC022031.2	3.37
DUXAP8	3.35	AC055854.1	2.10	LINC00858	4.84	FAM83A-AS1	3.46
IGFL2-AS1	4.02	AC009005.1	2.13	LINC02418	7.40	IGFL2-AS1	4.12
KIRC		LIHC		LUAD		PRAD	
lncRNA-RIDL	log2FC	lncRNA-RIDL	log2FC	lncRNA-RIDL	log2FC	lncRNA-RIDL	log2FC
PCED1B-AS1	2.85	LINC01018	-4.13	FENDRR	-4.56	MIR205HG	-2.20
AC011899.2	2.90	AL035661.1	-3.11	LHFPL3-AS2	-4.11	LINC01018	-1.99
AC087482.1	2.95	MAGI2-AS3	-2.85	LINC01936	-2.82	AP000808.1	-1.93
DARS-AS1	2.99	LINC02037	-2.63	SMIM25	-2.71	AF165147.1	-1.91

LINC00861	3.12	AC022784.6	-2.35	AC011899.2	-2.61	ADAMTS9-AS2	-1.83
SLC16A1-AS1	3.21	AC004862.1	-2.30	LINC00261	-2.60	LINC02562	-1.68
LINC00475	3.34	LINC02362	-2.27	AC116407.1	-2.53	AC036108.3	-1.57
AC019069.1	3.34	AC015908.3	-1.90	AC010329.1	-2.37	MAGI2-AS3	-1.56
AC026369.2	3.41	TMEM220-AS1	-1.88	AC125807.2	-2.26	AC013726.1	1.52
MIAT	3.54	LINC02499	-1.86	AC011511.5	-2.00	SNHG3	1.56
AC008105.3	3.66	HORMAD2-AS1	-1.83	LINC02362	2.13	LINC01297	1.58
LINC02188	3.78	LINC01485	-1.77	PVT1	2.42	AC004982.1	1.63
AL021328.1	3.81	LINC02535	-1.73	LUCAT1	2.47	AC008555.2	1.87
LINC00944	4.06	LINC00261	-1.67	LINC01270	2.55	FOXP4-AS1	2.25
HIF1A-AS2	4.08	HS1BP3-IT1	-1.65	AL354719.2	3.03	AL031123.2	2.36
AC019117.2	4.17	LINC00987	-1.52	LINC00511	3.06	DRAIC	2.39
PVT1	4.76	MIR99AHG	-1.51	AL391056.1	3.42	PCAT29	2.62
LUCAT1	5.00	AL021453.1	1.50	Z98257.1	4.73	AP000696.1	2.83
LINC00887	6.04	PVT1	1.69	AFAP1-AS1	5.24	AP004608.1	3.00
AL590644.1	6.29	LINC00511	1.98	FAM83A-AS1	5.32	AC009119.1	3.03

BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; KIRC: kidney renal clear cell carcinoma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; PRAD: prostate adenocarcinoma.

Given the importance of analyzing exclusive lncRNA-RIDLs genes for a cancer type, we observed that 81% of genes in the top 20 list appeared in just one type of cancer; i.e, only 30 of those genes are shared. An expression heatmap for the 130 distinct lncRNA-RIDLs genes is represented in Figure 2.

From shared lncRNA-RIDLs genes, we noticed that most present concordant expression direction among tumors, except for three of them: the AP004608.1, which is down-regulated in ESCA and up-regulated in PRAD; the AC011899.2, which appeared up-regulated in KIRC and down-regulated in LUAD; and the LINC02362, which appeared up-regulated in LUAD and down-regulated in LIHC (Table 3). As it could be expected, the BLCA and ESCA were the ones that shared the most lncRNA-RIDLs genes in the top 20: IGFL2-AS1; CADM3-AS1; and DUXAP8.

Top 20 IncRNA-RIDLs genes differentially expressed

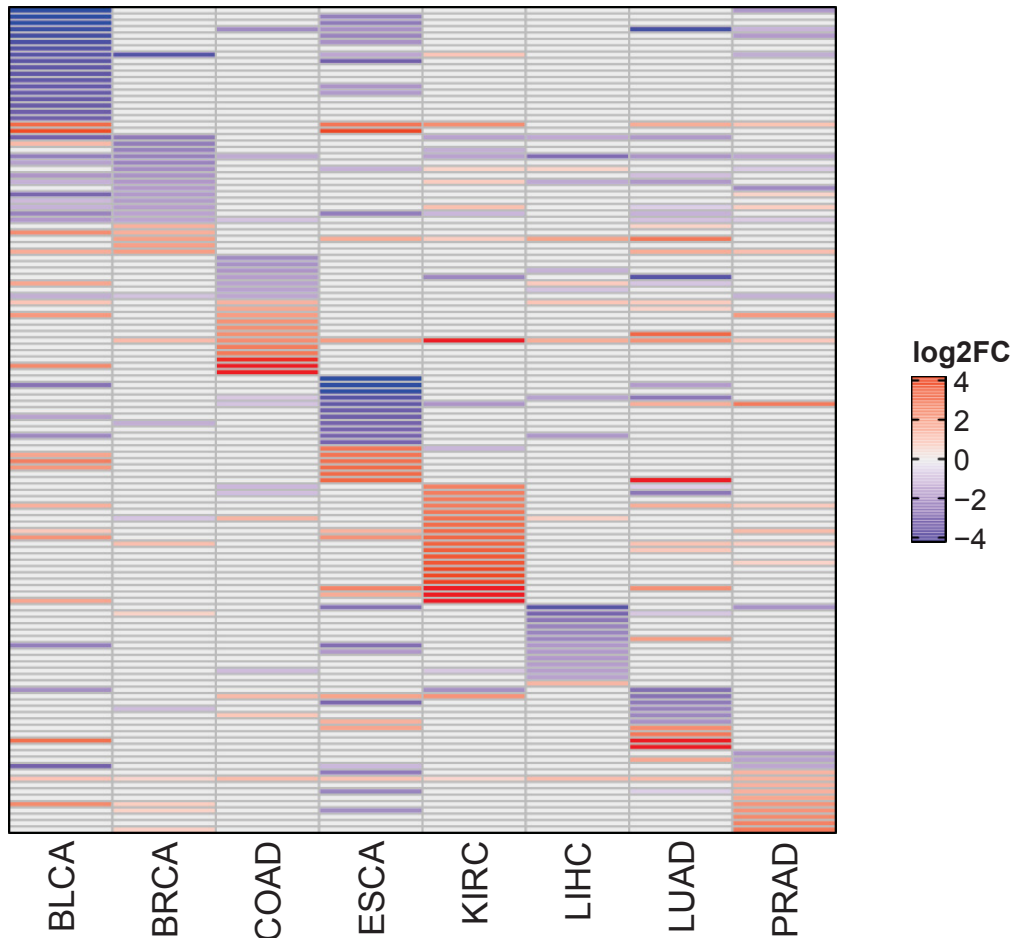


Figure 2. Heatmap of IncRNAs relative expression values comparing tumor samples and its non-tumoral counterpart. Rows represent IncRNA-RIDLs genes, and columns represent the analyzed eight cancer types. Expression values are represented as log₂fold change, varying between -5 and 5, with positive values represented in red and negative values represented in blue. BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; KIRC: kidney renal clear cell carcinoma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; PRAD: prostate adenocarcinoma.

Evaluation of TEs types in the top 20 deregulated IncRNA-RIDLs genes

In analyzing the 130 top 20 deregulated lncRNA-RIDLs genes, we identified a total of 206 RIDLs, since several of these lncRNAs contain more than one repeat (Supplementary Table 4). The gene AL031587.6, for example, has 10 RIDL fragments, and CARMN has 6 fragments, although most of the genes have less than three.

Reflecting that the human genome repeat fraction is mostly composed of retrotransposons, we observed only one RIDL derived from a DNA transposon. The majority of them are derived from SINE (150), followed by LINE (70) and LTR (20). The more abundant TE superfamily in the 130 lncRNAs is MIR, appearing 108 times.

Survival Analysis

Among 130 lncRNAs in top 20 lncRNA-RIDLs genes most differential, 44% (57) was associated with survival time (cox p-value <0.05) in at least one tumor type (Table 4).

Table 4: lncRNA-RIDLs genes associated with survival time in cancer

lncRNA-RIDLs	Survival Analysis (cox p-value < 0.05)							
	BLCA	BRCA	COAD	ESCA	KIRC	LIHC	LUAD	PRAD
AC135012.3	0.0361							
HAND2-AS1	0.0011							
FENRR							0.0094	
ADAMTS9-AS2	0.0030							
TARID	0.0287							
CARMN	0.0002	0.0322			0.0022			
FRMD6-AS2	0.0005							
AC134312.1	0.0299							
AC079313.2	0.0013							
MEF2C-AS1	0.0099							
AP003071.4	0.0352							
LINC01081	0.0241							
AP000911.1	0.0196							
DIO3OS	0.0157							

DUXAP8					0.0154
MIR99AHG	0.0001			0.0295	0.0001
SOX9-AS1	0.0263				
EMX2OS			0.0000		
MAGI2-AS3	0.0023				
LINC00987			0.0002		
PWAR6					0.0236
LINC00641					0.0214
AL021392.1					0.0420
LINC01977	0.0356				
LINC00511			0.0077	0.0004	0.0173
AC009005.1		0.0272			
LHFPL3-AS2					0.0086
LINC00654	0.0017	0.0361			
TMEM147-AS1	0.0021				0.0315
PVT1					0.0621
LINC00982			0.0451		0.0025
LINC00261		0.0054		0.0112	0.0108
AP004608.1					0.0140
AC015908.3				0.0101	0.0052
AC103702.2				0.0000	
LINC00460			0.0177		
AC087482.1				0.0009	
DARS-AS1	0.0161				0.0537
LINC00861				0.0491	
SLC16A1-AS1		0.0110			
AC026369.2				0.0024	
MIAT		0.0003			0.0077
LINC02188				0.0402	
AL021328.1				0.0180	
HIF1A-AS2				0.0024	
LINC00887			0.0456		
LINC01018			0.0348		
AC004862.1				0.0102	
LINC02362				0.0010	0.0087
TMEM220-AS1				0.0120	

LINC01485		0.0063
LINC02535		0.0408
LINC01936	0.0001	0.0198
AC010329.1		0.0402
Z98257.1	0.0019	
AC036108.3	0.0473	
SNHG3		0.0002
		0.0350

BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; KIRC: kidney renal clear cell carcinoma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; PRAD: prostate adenocarcinoma.

BLCA followed by LUAD were the cancer types with the highest number of lncRNAs associated with survival, with 23 and 15 lncRNA-RIDLs, respectively. The lncRNA-RIDLs related to survival in more cancer types were CARMN, LINC00261, LINC00511, and MIR99AHG. They appeared in three cancer types - BLCA, BRCA, KIRC; COAD, LIHC, LUAD; ESCA, KIRC, LIHC; and BLCA, LIHC, LUAD, respectively.

SNHG3, which appeared differentially expressed in the eight cancer types, was associated with survival in KIRC and LUAD.

lncRNA-RIDLs genes with a significant role in cancer

Recently, the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium has published the Cancer lncRNA Census (CLC), a compilation of GENCODE lncRNAs with causal roles in cancer phenotype whose first version contains lncRNA 122 genes²⁴. We observed that 19 of these genes (15.6%) have RIDLs and 12 correspond to lncRNAs present in the top 20 deregulate list in the eight types of cancer. Moreover, from these 12 lncRNA-RIDLs genes, we found seven (Table 5) with a significant survival rate in one or more cancers.

Table 5: LncRNA-RIDLs among the compilation of lncRNAs with causal roles in cancer phenotype.

IncrNA-RIDLs	Cancer type(s) with differentially expressed	Cancer type(s) with significant survival analysis	Oncogene (o) or Tumour Supp (t)	TE_name	TE_superfamily	TE_class/order
ADAMTS9-AS2	BLCA, ESCA, PRAD	BLCA	t	MIR3	MIR	SINE
				MIRb	MIR	SINE
FENDRR	BLCA, COAD, ESCA, LUAD, PRAD	LUAD	t	MIR3	MIR	SINE
				MIRb	MIR	SINE
LINC00261	COAD, ESCA, LIHC, LUAD	COAD, LIHC, LUAD	t	L2a	L2	LINE
LINC00511	BRCA, ESCA, KIRC, LIHC, LUAD	ESCA, KIRC, LIHC	o	MIR3	MIR	SINE
				MIRb	MIR	SINE
				MIRc	MIR	SINE
LINC00982	BLCA, ESCA, LUAD	ESCA, LUAD	t	MIRb	MIR	SINE
				MIRc	MIR	SINE
MIR99AHG	BLCA, BRCA, KIRC, LIHC, LUAD	BLCA, LIHC, LUAD	o	L2a	L2	LINE
				MER63A	hAT-Blackjack	DNA
PVT1	BRCA, COAD, ESCA, KIRC, LIHC, LUAD, PRAD	PRAD	o	MIRb	MIR	SINE

BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; KIRC: kidney renal clear cell carcinoma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; PRAD: prostate adenocarcinoma.

DISCUSSION

Identifying lncRNAs that play important putative functions within a myriad of them is a challenge. We focused on those lncRNA genes containing putative functional repeats, the RIDLs, derived from TEs. We found that 26.9% (944/3509) of differentially expressed lncRNAs in cancer contain predicted RIDLs, which underscores the importance of these lncRNAs in tumor cells.

Understanding domains derived from TEs found in lncRNA sequences can lead to a better understanding of lncRNA functions. A classic example of a

lncRNA carrying known functional RIDL is the XIST lncRNA, the structured A-repeat that is essential to recruiting the PRC2 silencing complex²⁶. In the present study, XIST was found deregulated in breast cancer samples, but this association is known for multiple tumor types, as well as it being a potential biomarker for solid tumor prognosis²⁷.

lncRNA FENDRR contains a putative DNA-binding domain derived from an L1 element¹⁹. In our study, FENDRR was among the top 20 most deregulated lncRNA genes, whereas Carlevaro-Fita did not predict this specific LINE insertion as RIDL²⁰. The RIDL highlighted in FENDRR contains MIR3 and MIRb, both SINEs, stressing the potential impact of TE insertions, still underexplored.

FENDRR has a known causal role in cancer and is described as abnormally expressed in various cancer types in the literature²⁸. We reinforce the deregulation of FENDRR in BLCA, COAD, ESCA, LUAD, and PRAD, as well as its association with time survival in LUAD.

Carlevaro-Fita et al., using RIDL elements, discovered a dose-dependent relationship between TEs and lncRNAs that seems to play a role in regulating the subcellular localization of lncRNAs²⁰. Alu elements promote nuclear retention of lncRNA through association with the nuclear matrix protein HNRNPK⁵. Nguyen et al. showed that this mechanism is responsible for the nuclear retention of the lncRNA MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) that contains an Alu insertion²⁹. We found high expression of MALAT1 in LIHC, COAD, and KIRC, which was expected since MALAT1 is overexpressed in a variety of cancer types²⁵.

TE insertions into lncRNA sequences can be essential function regulators; however, functional activity for most of these RIDLs was not confirmed experimentally. Further research is needed to confirm the status of these sequences as functional domains and their association with cancer and other diseases.

Regarding lncRNA-RIDLs, apparently more associated with cancer in general, three genes were differentially expressed in at least seven types: SNHG3, MAGI2-AS3, and PVT1.

The small nucleolar RNA host gene 3 (SNHG3) lncRNA was up-regulated in the eight cancer types; it is in the top 20 list of deregulated genes in PRAD and associated with poor survival in KIRC and LUAD. This lncRNA contains one predicted RIDL derived from an Alu element (SINE) (Supplementary Table 4). Overexpression of SNHG3 has been observed in several cancer types - breast cancer, lung cancer, colorectal cancer, hepatocellular carcinoma, renal cell carcinoma, prostate cancer, and bladder cancer³⁰. It is associated with increased proliferation, cell cycle progression, invasion, and metastasis of cancer cells³¹. Although this lncRNA-RIDL is not on the CLC list, SNHG3 is found in the nucleus and cytoplasm, where it participates in several mechanisms of action. In the nucleus, it acts on DNA methylation by regulating methylase, interacting with transcription factors and suppressing gene transcription. In the cytoplasm, SNHG3 acts as a miRNA sponge and inhibits translation³⁰. It is known that this lncRNA functions as a microRNA sponge for miR-330-5p in breast cancer cells and miR-515-5p in BLCA; it also has an oncogenic role in the progression of breast, lung, and bladder cancer^{32,33,34}. Overexpressed SNHG3 was significantly associated with poor survival and worse clinical outcomes, including clinical stage, histological grade, distant metastasis, and lymph node metastasis in human cancers³⁵.

The MAGI2 antisense RNA 3 (MAGI2-AS3) was down-regulated in BLCA, BRCA, COAD, KIRC, LIHC, LUAD, PRAD, and was associated with overall survival in BLCA. It contains two predicted RIDLs (L1 – LINE and MIR – SINE). It is known that these lncRNAs act as a sponge for miR-25, leading to an inhibition of cell invasion and migration in lung cancer. In breast cancer, MAGI2-AS3 plays an important role as a tumor suppressor by targeting Fas and FasL signaling and by regulating the miR-374a-PTEN axis^{36,37,38}. It also occurs as ceRNA, binding to miR-525-5p in ovarian cancer and acting as a tumor inhibitor³⁹. Our results showed that MAGI2-AS3 is a potential tumor suppressor.

PVT1, an oncogene located at locus 8q24, was up-regulated in seven of the eight cancer types (BRCA, COAD, ESCA, KIRC, LIHC, LUAD, PRAD); it was also associated with survival in PRAD, where its overexpression was related to a worse prognosis. PVT1 is in the list of lncRNAs with a causal role in cancer according to the CLC. It contains one predicted RIDL (MIR - SINE) and has a close functional relationship with myelocytomatosis (Myc), which has oncogenic effects in several cancers⁴⁰. Its overexpression was previously associated with various cancer types and was related to poor prognosis⁴¹. PVT1 acts as a sponge for microRNA-145, inhibiting cell migration and invasion and promoting cell apoptosis in esophageal carcinoma⁴².

Many lncRNA-RIDLs genes are differentially expressed in only one cancer type. They corresponded to 62.71% (592) of all differentially expressed lncRNA-RIDLs (944) ($P < 0.01$) in the eight cancers analyzed. In the heatmap of the relative expression values of lncRNAs comparing tumor samples and their non-tumoral counterparts (Figure 2), the potential of top-20 lncRNA-RIDLs in different tumor types is evident.

This specificity is also found in lncRNAs in general⁴, highlighting the potential use of these molecules as biomarkers. Additionally, specific TE domains may be significantly related to tissue-specific expression levels of lncRNAs. In 2018, Chishima, Iwakiri and Hamada⁴³ identified TEs that were significantly related to tissue-specificity. Many TE-tissue pairs were detected, indicating that multiple TE families can be functional domains or regulatory sequences of specific tissue expression⁴³. In cancer cells, this aspect was not investigated; but, based on the great number of specific lncRNA-RIDLs differentially expressed in exclusive tumor types, it is plausible to assume the influence of this domain on expression regulation carcinogenesis.

In addition to the differential expression in patients with cancer, an essential parameter of vast clinical application is the association between expression level and patient survival time. Among lncRNAs in the top 20 lncRNA-RIDL genes, the most differentially expressed ones were associated with survival time in at least

one tumor type (44%, 57/130), showing the great added potential of lncRNA-RIDLs as prognostic markers.

Most lncRNAs are still poorly understood; however, the CLC list describes 122 lncRNAs that are well-characterized and play causal roles in cancer phenotypes. Among them, 19 (15.6%) have RIDLs.

Our study described lncRNAs-RIDLs differentially expressed in cancer, associated with time survival of patients and with a recognized causal role in tumors. Among the molecules highlighted in the three analyses, we found seven lncRNA-RIDLs, including FENDRR, PVT1, ADAMTS9-AS2, MIR99AHG, LINC00511, LINC00982, and LINC00261. Among these seven lncRNA-RIDLs, five of them have miR/SINE transposable elements, including FENDRR (with MIR3 and miRb) and PVT1 (with miRb TEs), as previously discussed.

ADAMTS9-AS2, with MIR3 and miRb RIDLs, presented tumor suppression activity. Our results reinforced its hypoexpression in BLCA, ESCA, and PRAD, as well as its association with survival time in BLCA. The CLC acknowledges the causal effect of ADAMTS9-AS2 in cancer, which is related to the proliferation, invasion, migration, and inhibition of apoptosis through various ways, such as regulating miRNAs and activating classical signaling pathways, like PI3K / Akt / mTOR⁴⁴.

LINC00511, an oncogene located at chromosome 17, was up-regulated in BRCA, ESCA, KIRC, LIHC, and LUAD and was associated with survival in ESCA, KIRC and LIHC. It is correlated to poor prognosis in triple-negative breast cancer, LIHC, and LUAD, accelerating proliferation and migration, and thus aggravating tumor progression^{45,46,47}. In LUAD and KIRC, linc00511 acts as a sponge for miRNA - miRNA-625-5p and miRNA-625, respectively^{48,47}.

In addition to containing SINE elements such as MIR3, MIRb and MIRc, LINC00982 was deregulated in BRCA, ESCA, KIRC, LIHC, and LUAD, and associated with survival time in ESCA, KIRC, and LIHC. According to the literature, LINC00982 regulates cell proliferation, having clinical relevance in

patients with gastric (PMID: 26334618), lung (PMID: 30132554), thyroid, and renal carcinomas, mainly influencing PI3K-AKT signaling pathways^{49,50}.

Considering all cancer types, SINE elements were also the most abundant TEs among the RIDLs, mostly from the MIR (Mammalian-wide interspersed repeats) superfamily. MIRs are ancient elements of tRNA-derived SINEs with more than 500,000 copies in the human genome and constitute a significant fraction of the core of genic enhancers that show strong association levels of gene expression, tissue-specific gene expression, and cellular functions⁵¹.

Additionally, MIR elements were predicted to play the role of insulators in the human genome. Insulator regulatory elements help to organize chromatin by creating barriers and inducing silencing. These predicted MIR insulators were computationally tested to serve as chromatin barriers and regulators of gene expression in T cells⁵², including in intergenic and intronic genome regions. However, RIDLs listed by Carvelaro-Fita et al. (2019) focus on exons containing different domains; thus, the role of MIRs in RNA sequences needs to be better investigated for potential clues about lncRNAs mechanisms of action.

The second most abundant element among RIDLs are L2 superfamily elements (LINE). LINC00261 contains a L2a element and was deregulated in COAD, ESCA, LIHC, and LUAD, as well as further associated with survival time in COAD, LIHC, and LUAD.

LINC00261 was deeply studied as being deregulated in multiple tumor types and as regulating proliferation, migration, invasion, and the formation of metastasis^{53,54}. The cell mechanism of LINC00261 seems to be intricately associated with nuclear and cytoplasmic functions. Recently, a competitive endogenous RNA network that sequesters miR-8485 from targeting the CBX2 mRNA was described in prostate cancer; while in the nucleus, LINC00261 acts as a scaffold to induce SMAD-driven expression of the FOXA2 gene⁵⁵.

LINC00261 regulated important pathways, including Notch signaling⁵⁴, Wnt/ β -catenin pathway^{56,57} and regulating Slug⁵⁸, and c-myc proteins⁵⁹. It is also associated with many competitive endogenous RNA network^{60,61,62,63}.

Interestingly, CAO et al. showed that L2 elements, the second most abundant TE in the 206 RIDLs, also act as enhancers in different human cell types⁶⁴. Moreover, these authors found that MIR and L2 elements can have promoter activity as well, and frequently share long-range intra-chromosomal interactions and binding of physically interacting transcription factors. Just as MIR, L2 elements are also ancient and have the same proportion of copies in the genome⁶⁴.

Less frequent in RIDLs are DNA TE elements. MiR99AHG contains a L2a RIDL (LINE) and also a DNA element, MER63A (superfamily hAT-Blackjack). MiR99AHG was deregulated in BLCA, BRCA, KIRC, LIHC, and LUAD and associated with survival time in BLCA, LIHC, and LUAD. Previously, MIR99AHG expression was associated with overall survival in gastric⁶⁵, head and neck⁶⁶, and lung cancer⁶⁷. Additionally, it has been suggested to act in competitive endogenous RNA^{65,67}.

The role of hAT-Blackjack in MIR99AHG has not been previously investigated. Some suggest that the DNA.hAT.Blackjack family is associated with a more specific expression in lung tissues based on the hypothesis that TEs may affect tissue specificity of lncRNA expression. LUAD was prominent in our analysis; MIR99HG was deregulated and related to overall survival in lung cancer patients⁴³.

The diversity of lncRNA-RIDL genes in terms of length, number, structure, distribution, and function renders them a poorly understood class with great potential for investigations. Despite recent studies showing the importance of lncRNAs containing RIDLs in cell biology, their deregulation in cancer has not been deeply studied.

We analyzed in detail the expression of lncRNA-RIDLs in eight cancer types and identified 944 differentially expressed lncRNA-RIDLs by comparing them to their non-tumoral counterparts. Analysis of the differential expression of all lncRNA-RIDLs revealed that most of these molecules were concordant as to their dysregulation in different tumor types. 62.71% of them appeared in only one of the eight cancer types. These molecules were also able to differentiate tumor types and play an important role in each cellular context.

The three lncRNA-RIDL genes that stood out, present in at least seven of the cancer types under study, were SNHG3, MAGI2-AS3, and PVT1; their RIDLs were identified as FRAM, L1MC4, and MIRb, respectively. In the list of the top 20 lncRNA-RIDLs in the eight cancer types (130 differentially expressed lncRNA-RIDLs), the most abundant TE class/order was SINE, and the TE superfamily that appeared most often was MIR.

In the list of 130 differentially expressed lncRNA-RIDLs, 44% (57) were associated with survival time in at least one cancer type. The lncRNA-RIDL genes related to survival in most cancer types were CARMN, LINC00261, LINC00511, and MIR99AHG. Additionally, among a compilation of 122 lncRNAs with causal roles in cancer phenotype, 19 (15.6%) have RIDLs, and 12 correspond to lncRNA-RIDLs present in the top 20 lncRNA-RIDL genes list.

Regarding the molecules highlighted in differentially expressed genes associated with survival time of patients and a recognized causal role in tumors, we found seven lncRNA-RIDLs, including FENDRR, PVT1, ADAMTS9-AS2, LINC00511, LINC00982, MIR99AHG, and LINC00261.

More studies are necessary to characterize and validate the highlighted lncRNA-RIDLs genes and to better understand the function of these sequences in lncRNA roles. Our study contributed to the understanding of lncRNA-RIDLs genes and indicated new possibilities in the identification biomarkers, emphasizing the importance of these molecules in cancer.

ACKNOWLEDGMENTS

CAPES, Fundação Araucária, and CNPq to funding.

Supplementary table 1 – enviado em arquivo de Excel.

Supplementary Table 2 - enviado em arquivo de Excel.

Supplementary table 3 - enviado em arquivo de Excel.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA CANCER J CLIN.* 2021;0(0):1-41.
2. Hanahan D, Weinberg RA. Review Hallmarks of Cancer : The Next Generation. *Cell.* 2011;144(5):646-674.
3. Anastasiadou E, Jacob LS, Slack FJ, Israel B. Non-coding RNA networks in cancer. *Nat Rev Cancer.* 2019;18(1):5-18.
4. Oliveira JC de, Oliveira LC, Mathias C, et al. Long non - coding RNAs in cancer : Another layer of complexity. *Wiley.* 2019;(December 2018):1-18.
5. Yao R, Wang Y, Chen L. Cellular functions of long noncoding RNAs. *Nat Cell Biol.* 2019;21(May).
6. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol.* 2013;14(11):699-712.
7. Fang Y, Fullwood MJ. Roles , Functions , and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics Proteomics Bioinformatics.* 2016;14(1):42-54.
8. Hon C, Ramilowski JA, Harshbarger J, et al. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature.* 2019;543(7644):199-204.
9. Hadjiargyrou M, Delihias N. The Intertwining of Transposable Elements and Rnas, Non-coding. *Int J Mol Sci.* 2013:13307-13328.
10. Kapusta A, Kronenberg Z, Lynch VJ, et al. Transposable Elements Are Major Contributors to the Origin , Diversification , and Regulation of Vertebrate Long Noncoding RNAs. *PLOS Genet.* 2013;9(4).
11. Human I, Sequencing G. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(February).
12. Bourque G, Burns KH, Gehring M, et al. Ten things you should know

- about transposable elements. *Genome Biol.* 2018:1-12.
13. Bennett EA, Coleman LE, Tsui C, Pittard WS, Devine SE. Natural Genetic Variation Caused by Transposable Elements in Humans. *Genet Soc Am.* 2004;951(October):933-951.
 14. Bodega B, Orlando V. ScienceDirect Repetitive elements dynamics in cell identity programming, maintenance and disease. *Curr Opin Cell Biol.* 2014;31:67-73.
 15. Abascal F, Tress ML, Valencia A. Sequence analysis Alternative splicing and co-option of transposable elements : the case of TMPO / LAP2 a and ZNF451 in mammals. *Bioinformatics.* 2015;31(March):2257-2261.
 16. Criscione SW, Zhang Y, Thompson W, Sedivy JM, Neretti N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics.* 2014:1-17.
 17. Etchegaray E, Naville M, Volff J, Haftek-terreau Z. Transposable element-derived sequences in vertebrate development. *Mob DNA.* 2021:1-24.
 18. Mager DL, Lagemaat LN Van De, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Genome Anal Transposable.* 2003;19(10):530-536.
 19. Johnson R, Guigó R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA.* 2014;20:959-976.
 20. Carlevaro-fita J, Polidori T, Das M, Navarro C, Zoller TI, Johnson R. Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. *Genome Res.* 2019:208-222.
 21. Li R, Qu H, Wang S, et al. GDCRNATools : an R / Bioconductor package for integrative analysis of lncRNA , miRNA , and mRNA data in GDC. *Bioinformatics.* 2018;(March):2-4.
 22. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7).
 23. Harrow J, Frankish A, Gonzalez JM, Frazer KA. GENCODE : The reference human genome annotation for The ENCODE Project. *Genome Res.* 2012:1760-1774.
 24. Gu Z, Eils R, Schlesner M. Genome analysis Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Genome Anal.* 2016:20-22.
 25. Carlevaro-fita J, Lanzós A, Feuerbach L, et al. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs. *Commun Biol.* 2020;(1mim):1-16.
 26. Elisaphenko EA, Kolesnikov NN, Shevchenko AI, et al. A Dual Origin of the Xist Gene from a Protein-Coding Gene and a Set of Transposable

- Elements. *PLoS One*. 2008;3(6):1-11.
27. Zhou Q, Hu W, Zhu W, et al. Clinica Chimica Acta Long non coding RNA XIST as a prognostic cancer marker – A meta-analysis. *Clin Chim Acta*. 2018;482(January):1-7.
 28. Zheng Q, Zhang Q, Yu X, He Y, Guo W. FENDRR: A pivotal, cancer-related, long non-coding RNA Qingyuan. *Biomed Pharmacother*. 2021;137(February):111390.
 29. Nguyen TM, Kabotyanski EB, Reineke LC, et al. The SINEB1 element in the long non-coding RNA Malat1 is necessary for TDP-43 proteostasis. *Nucleic Acids Res*. 2020;48(5):2621-2642.
 30. Xu B, Mei J, Ji W, et al. LncRNA SNHG3 , a potential oncogene in human cancers. *Cancer Cell Int*. 2020:1-11.
 31. Zimta A, Tigu AB, Braicu C, Stefan C. An Emerging Class of Long Non-coding RNA With Oncogenic Role Arises From the snoRNA Host Genes. *Front Oncol*. 2020;10(April).
 32. Li Y, Zhao Z, Liu W, Li X. SNHG3 Functions as miRNA Sponge to Promote Breast Cancer Cells Growth Through the Metabolic Reprogramming. *Appl Biochem Biotechnol*. 2020.
 33. Dai G, Huang C, Yang J, et al. LncRNA SNHG3 promotes bladder cancer proliferation and metastasis through miR-515-5p / GINS2 axis. *J Cell Mol Med*. 2020;3(May):9231-9243.
 34. Liu L, Ni J, He X. Upregulation of the Long Noncoding RNA SNHG3 Promotes Lung Adenocarcinoma Proliferation. *Dis Markers*. 2018;2018.
 35. Jiang Y, Le L. Overexpression of lncRNA SNGH3 Predicts Unfavorable Prognosis and Clinical Outcomes in Human Cancers: Evidence from a Meta-Analysis. *Biomed Res Int*. 2020;2020:13.
 36. Yang Y, Yang H, Xu M, et al. Long non - coding RNA (lncRNA) MAGI2 - AS3 inhibits breast cancer cell growth by targeting the Fas / FasL signalling pathway. *Hum Cell*. 2018;(0123456789).
 37. Du S, Hu W, Zhao Y, et al. Long non-coding RNA MAGI2-AS3 inhibits cell proliferation and invasion. *Cancer Biomarkers*. 2019;1:1-9.
 38. Sui Y, Chi W, Feng L, Jiang J. LncRNA MAGI2-AS3 is downregulated in non-small cell lung cancer and may be a sponge of miR-25. *BMC Pulm Med*. 2020:1-7.
 39. Chang H. MAGI2-AS3 suppresses MYC signaling to inhibit cell proliferation and migration in ovarian cancer through targeting. *Cancer Med*. 2020;(December 2019):6377-6386.
 40. Li Y, Egranov SD, Yang L, Lin C. Molecular mechanisms of long noncoding RNAs-mediated cancer metastasis. *Wiley Period*. 2019;58(July 2018):200-207.

41. Lu D, Luo P, Wang Q, Ye Y, Wang B. Clinica Chimica Acta lncRNA PVT1 in cancer : A review and meta-analysis. *Clin Chim Acta*. 2017;474(September):1-7.
42. Shen S. Down-regulation of long noncoding RNA PVT1 inhibits esophageal carcinoma cell migration and invasion and promotes cell apoptosis via microRNA-145-mediated inhibition of FSCN1. *Mol Oncol*. 2019;13:2554-2573.
43. Chishima T, Iwakiri J, Hamada M. Identification of Transposable Elements Contributing to Tissue-Specific Expression of Long Non-Coding RNAs. *Genes (Basel)*. 2018;9(23).
44. Xu W, Wang B, Cai Y, et al. ADAMTS9-AS2: A functional long non-coding RNA in tumorigenesis. *Curr Pharm Des*. 2021.
45. Hu P, Cui H, Lei T, Li S, Mai E, Jia F. Linc00511 Indicates A Poor Prognosis Of Liver Hepatocellular Carcinoma. *OncoTargets Ther* 201912. 2019;9367-9376.
46. Liu R, Wang L, Gan T. Long noncoding RNA LINC00511 promotes cell growth and invasion in triple-negative breast cancer by interacting with Snail. *Cancer Manag Res*. 2019;11:5691–5699.
47. Xue J. LncRNA LINC00511 plays an oncogenic role in lung adenocarcinoma by regulating PKM2 expression via sponging miR-625-5p. *Thorac Cancer*. 2020;11:2570-2579.
48. Deng H, Huang C, Wang Y, Jiang H, Peng S, Zhao X. LINC00511 promotes the malignant phenotype of clear cell renal cell carcinoma by sponging microRNA-625 and thereby increasing cyclin D1 expression. *Aging (Albany NY)*. 2019;11(16).
49. Xu D, Yu J, Zhuang S, Zhang S, Hong Z, Yuan C. Overexpression of long non-coding RNA LINC00982 suppresses cell proliferation and tumor growth of papillary thyroid carcinoma through PI3K-ATK signaling pathway. *Portl Press*. 2019;0(July):1-10.
50. Zhang C, Li X, Luo Z, Wu T, Hu H. Upregulation of LINC00982 inhibits cell proliferation and promotes cell apoptosis by regulating the activity of PI3K / AKT signaling pathway in renal cancer. *Eur Rev Med Pharmacol Sci*. 2019;23:1443-1450.
51. Jjingo D, Conley AB, Wang J, Mariño-ramírez L, Lunyak V V, Jordan IK. Mammalian-wide interspersed repeat (MIR) -derived enhancers and the regulation of human gene expression. *Mob DNA*. 2014:1-12.
52. Wang J, Vicente-garcía C, Seruggia D, Moltó E, Fernandez-miñán A, Neto A. MIR retrotransposon sequences provide insulators to the human genome. *PNAS*. 2015:1-10.
53. Chen Z, Xiang L, Hu Z, et al. Epigenetically silenced linc00261 contributes to the metastasis of hepatocellular carcinoma via inducing the

- deficiency of FOXA2 transcription. *Am J Cancer Res.* 2021;11(1):277-296.
54. Zhang H, Li W, Han Y. LINC00261 suppresses cell proliferation, invasion and Notch signaling pathway in hepatocellular carcinoma. *Cancer Biomarkers.* 2017;1:1-8.
 55. Mather RL, Parolia A, Carson SE, et al. The evolutionarily conserved long non-coding RNA LINC00261 drives neuroendocrine prostate cancer proliferation and metastasis via distinct nuclear and cytoplasmic mechanisms. *Mol Oncol.* 2021;(2020):0-3.
 56. Shi J, Ma H, Wang H, et al. Overexpression of LINC00261 inhibits non – small cell lung cancer cells progression by interacting with miR - 522 - 3p and suppressing Wnt signaling. *J Cell Biochem.* 2019;(May):1-10.
 57. Yan D, Liu W, Liu Y, Luo M. LINC00261 suppresses human colon cancer progression via sponging miR - 324 - 3p and inactivating the Wnt / β - catenin pathway. *Cell Physiol.* 2019;(January):1-9.
 58. Yu Y, Zheng Z, Chen S, Chen E, Hu Y. Long non-coding RNA linc00261 suppresses gastric cancer progression via promoting Slug degradation. 2017; 21(5):955-967.
 59. Liu S, Zheng Y, Zhang Y, et al. Theranostics Methylation-mediated LINC00261 suppresses pancreatic cancer progression by epigenetically inhibiting c-Myc transcription. 2020:10(23).
 0. Zhu J, Deng J, Zhang L, et al. Reconstruction of lncRNA-miRNA-mRNA network based on competitive endogenous RNA reveals functional lncRNAs in skin cutaneous melanoma. 2020:1-20.
 61. Wang Z, Zhang J, Yang B, et al. Long Intergenic Noncoding RNA 00261 Acts as a Tumor Suppressor in Non-Small Cell Lung Cancer via Regulating miR-105 / FHL1 Axis. *J Cancer.* 2019;10.
 62. Fang Q, Sang L, Du S. Long noncoding RNA LINC00261 regulates endometrial carcinoma progression by modulating miRNA / FOXO1 expression. 2018;(June):1-8.
 63. Li Y. LINC00261 / microRNA-550a-3p / SDPR axis affects the biological characteristics of breast cancer stem cells. 2020;(June):1-14.
 64. Cao Y, Chen G, Wu G, et al. Widespread roles of enhancer-like transposable elements in cell identity and long-range genomic interactions. *Genome Res.* 2019:40-52.
 65. Meng Q, Wang X, Xue T, Zhao Q, Wang W, Zhao K. Long noncoding RNA MIR99AHG promotes gastric cancer progression by inducing EMT and inhibiting apoptosis via miR577 / FOXP1 axis. *Cancer Cell Int.* 2020:1-10.
 66. Yang B, Shen J, Xu L, Chen Y, Che X, Qu X. Genome-Wide Identification

of a Novel Eight-lncRNA Signature to Improve Prognostic Prediction in Head and Neck Squamous Cell Carcinoma. *Front Oncol.* 2019;9(September).

67. Ning P, Wu Z, Hu A, et al. Integrated genomic analyses of lung squamous cell carcinoma for identification of a possible competitive endogenous RNA network by means of TCGA datasets. *PeerJ.* 2018:1-18.

6 CONCLUSÃO

- ✓ Nos oito tipos de câncer analisados (BLCA, BRCA, COAD, ESCA, COAD, KIRC, LIHC, LUAD and PRAD), foram identificados 944 genes lncRNA-RIDLs expressos diferencialmente em comparação com sua contraparte não tumoral;
- ✓ 62,71% dos lncRNA-RIDLs aparecem em apenas um dos oito tipos de câncer. Essas moléculas também foram capazes de diferenciar os tipos de tumor;
- ✓ Três genes lncRNA-RIDLs se destacaram por estarem presentes em pelo menos sete tipos de câncer: SNHG3, MAGI2-AS3 e PVT1. As sequências destes RIDLs são: FRAM, L1MC4 e MIRb, respectivamente;
- ✓ Ao analisar todos os genes lncRNA-RIDLs da lista dos 20 principais lncRNA-RIDLs nos oito tipos de câncer (130 lncRNA-RIDLs expressos diferencialmente), a classe/ordem TE mais abundante foi SINE e a superfamília de TE foi MIR;
- ✓ Na lista dos 130 lncRNA-RIDLs expressos diferencialmente, 44% (57) foi associado ao tempo de sobrevida em pelo menos um tipo de tumor. Os lncRNA-RIDLs relacionados à sobrevida em mais tipos de câncer foram: CARMN, LINC00261, LINC00511 e MIR99AHG; e
- ✓ Entre uma compilação de 122 lncRNAs com papéis causais no fenótipo do câncer, 19 desses genes (15,6%) têm RIDLs e 12 correspondem aos lncRNA-RIDLs presentes na lista dos 20 principais lncRNA-RIDLs expressos diferencialmente nos oito tipos de câncer.

Mais estudos são necessários para caracterizar e validar o destaque dos genes lncRNA-RIDLs e entender melhor a função do RIDL nos lncRNAs. Nosso estudo trouxe conhecimento adicional sobre os genes lncRNA-RIDLs e novas possibilidades na identificação de biomarcadores, destacando a importância dessas moléculas no câncer.

REFERÊNCIAS

- ABASCAL, F.; TRESS, M. L.; VALENCIA, A. Sequence analysis Alternative splicing and co-option of transposable elements : the case of TMPO / LAP2 a and ZNF451 in mammals. **Bioinformatics**, v. 31, n. March, p. 2257–2261, 2015.
- ANASTASIADOU, E.; JACOB, L. S.; SLACK, F. J.; ISRAEL, B. Non-coding RNA networks in cancer. **Nat Rev Cancer**, v. 18, n. 1, p. 5–18, 2019.
- BAO, W.; KOJIMA, K. K.; KOHANY, O. Repbase Update , a database of repetitive elements in eukaryotic genomes. **Mobile DNA**, p. 4–9, 2015. Mobile DNA. Disponível em: <<http://dx.doi.org/10.1186/s13100-015-0041-9>>. .
- BENNETT, E. A.; COLEMAN, L. E.; TSUI, C.; PITTARD, W. S.; DEVINE, S. E. Natural Genetic Variation Caused by Transposable Elements in Humans. **Genetics Society of America**, v. 951, n. October, p. 933–951, 2004.
- BHAN, A.; SOLEIMANI, M.; MANDAL, S. S. Long Noncoding RNA and Cancer : A New Paradigm. **Cancer Research**, p. 1–18, 2017.
- BIÉMONT, C.; VIEIRA, C. NEWS & VIEWS FEATURE Junk DNA as an evolutionary force. **Nature**, v. 443, n. October, p. 521–524, 2006.
- BODEGA, B.; ORLANDO, V. ScienceDirect Repetitive elements dynamics in cell identity programming , maintenance and disease. **Current Opinion in Cell Biology**, v. 31, p. 67–73, 2014. Elsevier Ltd. Disponível em: <<http://dx.doi.org/10.1016/j.ceb.2014.09.002>>. .
- BOURQUE, G.; BURNS, K. H.; GEHRING, M.; et al. Ten things you should know about transposable elements. **Genome Biology**, p. 1–12, 2018. Genome Biology.
- BRAY, F.; FERLAY, J.; SOERJOMATARAM, I. Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA CANCER J CLIN**, p. 394–424, 2018.
- CARLEVARO-FITA, J.; LANZÓS, A.; FEUERBACH, L.; et al. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs. **Communications Biology**, , n. Imim, p. 1–16, 2020.
- CARLEVARO-FITA, J.; POLIDORI, T.; DAS, M.; et al. Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. **Genome Research**, p. 208–222, 2019.
- CHEN, L. Linking Long Noncoding RNA Localization and Function. **Trends in Biochemical Sciences**, v. 41, n. 9, p. 761–772, 2016. Elsevier Ltd. Disponível em: <<http://dx.doi.org/10.1016/j.tibs.2016.07.003>>. .
- CRISCIONE, S. W.; ZHANG, Y.; THOMPSON, W.; SEDIVY, J. M.; NERETTI, N. Transcriptional landscape of repetitive elements in normal and cancer human cells. **BMC Genomics**, p. 1–17, 2014.

- CTCA, C. T. C. OF A. No Title. Disponível em: <<https://www.cancercenter.com/cancer-types/kidney-cancer/types>>. Acesso em: 3/10/2021.
- ESTELLER, M.; PANDOLFI, P. P.; SCIENCES, H.; et al. The Epitranscriptome of Non-coding RNAs in Cancer. **Cancer Discov.**, v. 7, n. 4, p. 359–368, 2018.
- ETCHEGARAY, E.; NAVILLE, M.; VOLFF, J.; HAFTEK-TERREAU, Z. Transposable element-derived sequences in vertebrate development. **Mobile DNA**, p. 1–24, 2021. Mobile DNA.
- FANG, Y.; FULLWOOD, M. J. Roles , Functions , and Mechanisms of Long Non-coding RNAs in Cancer. **Genomics, Proteomics & Bioinformatics**, v. 14, n. 1, p. 42–54, 2016. Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China. Disponível em: <<http://dx.doi.org/10.1016/j.gpb.2015.09.006>>. .
- FERRÉ, F.; COLANTONI, A.; HELMER, M. Revealing protein – lncRNA interaction Fabrizio Ferre. **OXFORD**, v. 17, n. April 2015, p. 106–116, 2016.
- FINNEGAN, D. J. ELEMENTS and GENOME EVOLUTION. **Reviws**, v. 5, n. 4, 1989.
- GEISLER, S.; COLLER, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. **Nature Reviews Molecular Cell Biology**, v. 14, n. 11, p. 699–712, 2013.
- GU, Z.; EILS, R.; SCHLESNER, M. Genome analysis Complex heatmaps reveal patterns and corre- lations in multidimensional genomic data. **Genome analysis**, p. 20–22, 2016.
- HADJIARGYROU, M.; DELIHAS, N. The Intertwining of Transposable Elements and Rnas, Non-coding. **International Journal of Molecular Sciences**, p. 13307–13328, 2013.
- HANAHAN, D.; WEINBERG, R. A. Review Hallmarks of Cancer : The Next Generation. **Cell**, v. 144, n. 5, p. 646–674, 2011. Elsevier Inc. Disponível em: <<http://dx.doi.org/10.1016/j.cell.2011.02.013>>. .
- HARROW, J.; FRANKISH, A.; GONZALEZ, J. M.; FRAZER, K. A. GENCODE : The reference human genome annotation for The ENCODE Project. **Genome Res.**, p. 1760–1774, 2012.
- HON, C.; RAMIŁOWSKI, J. A.; HARSHBARGER, J.; et al. An atlas of human long non-coding RNAs with accurate 5' ends. **Nature**, v. 543, n. 7644, p. 199–204, 2019.
- HUMAN, I.; SEQUENCING, G. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. February, 2001.
- INCA, I. N. DE C. No Title. Disponível em: <<https://www.inca.gov.br/numeros-de-cancer>>. Acesso em: 3/10/2020.
- JOHNSON, R.; GUIGÓ, R. The RIDL hypothesis : transposable elements as

functional domains of long noncoding RNAs. **RNA**, v. 20, p. 959–976, 2014.

KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. , v. 181, n. 2005, p. 2006–2007, 2007.

KAPUSTA, A.; KRONENBERG, Z.; LYNCH, V. J.; et al. Transposable Elements Are Major Contributors to the Origin , Diversification , and Regulation of Vertebrate Long Noncoding RNAs. **PLOS Genetics**, v. 9, n. 4, 2013.

KWOK, Z. H.; TAY, Y. Long noncoding RNAs : links between human health and disease. **Portland Press**, v. 45, p. 805–812, 2017.

LEVIN, H. L.; MORAN, J. V.; SHRIVER, E. K. Dynamic interactions between transposable elements and their hosts. **Nature Publishing Group**, v. 12, n. 9, p. 615–627, 2011. Nature Publishing Group. Disponível em: <<http://dx.doi.org/10.1038/nrg3030>>. .

LI, L.; LENG, R.; FAN, Y.; PAN, H.; YE, D. Author ' s Accepted Manuscript. **Experimental Cell Research**, 2017. Disponível em: <<https://doi.org/10.1016/j.yexcr.2017.10.010>>. .

LI, R.; QU, H.; WANG, S.; et al. GDCRNATools : an R / Bioconductor package for integrative analysis of lncRNA , miRNA , and mRNA data in GDC. **Bioinformatics**, , n. March, p. 2–4, 2018.

LI, Y.; EGRANOV, S. D.; YANG, L.; LIN, C. Molecular mechanisms of long noncoding RNAs-mediated cancer metastasis. **Wiley Periodicals**, v. 58, n. July 2018, p. 200–207, 2019.

LIU, L.; NI, J.; HE, X. Upregulation of the Long Noncoding RNA SNHG3 Promotes Lung Adenocarcinoma Proliferation. **Disease Markers**, v. 2018, 2018. Hindawi.

MAGER, D. L.; LAGEMAAT, L. N. VAN DE; MEDSTRAND, P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. **Genome Analysis Transposable**, v. 19, n. 10, p. 530–536, 2003.

NGUYEN, T. M.; KABOTYANSKI, E. B.; REINEKE, L. C.; et al. The SINEB1 element in the long non-coding RNA Malat1 is necessary for TDP-43 proteostasis. **Nucleic Acids Research**, v. 48, n. 5, p. 2621–2642, 2020.

OLIVEIRA, J. C. DE; OLIVEIRA, L. C.; MATHIAS, C.; et al. Long non - coding RNAs in cancer : Another layer of complexity. **Wiley**, , n. December 2018, p. 1–18, 2019.

PARASRAMKA, M. A.; MAJI, S.; MATSUDA, A.; YAN, I. K.; PATEL, T. Pharmacology & Therapeutics Long non-coding RNAs as novel targets for therapy in hepatocellular carcinoma. **Pharmacology and Therapeutics**, v. 161, p. 67–78, 2016. The Authors. Disponível em: <<http://dx.doi.org/10.1016/j.pharmthera.2016.03.004>>. .

- RITCHIE, M. E.; PHIPSON, B.; WU, D.; et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. **Nucleic Acids Research**, v. 43, n. 7, 2015.
- SUNG, H.; FERLAY, J.; SIEGEL, R. L.; et al. Global Cancer Statistics 2020 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA CANCER J CLIN**, v. 0, n. 0, p. 1–41, 2021.
- TONG, J.; MA, X.; YU, H.; YANG, J. SNHG15 : a promising cancer-related long noncoding RNA. **Cancer Management and Research**, v. 11, 2019.
- WOGAN, G. N.; HECHT, S. S.; FELTON, J. S.; CONNEY, A. H.; LOEB, L. A. Environmental and chemical carcinogenesis. **Seminars in Cancer Biology**, v. 14, p. 473–486, 2004.
- WU, T.; DU, Y. LncRNAs: From basic research to medical application. **International Journal of Biological Sciences**, v. 13, n. 3, p. 295–307, 2017.
- YANG, G.; LU, X.; YUAN, L. LncRNA: A link between RNA and cancer. **BBA - Gene Regulatory Mechanisms**, v. 1839, n. 11, p. 1097–1109, 2014.
- YAO, R.; WANG, Y.; CHEN, L. Cellular functions of long noncoding RNAs. **Nature Cell Biology**, v. 21, n. May, 2019. Springer US. Disponível em: <<http://dx.doi.org/10.1038/s41556-019-0311-8>>. .
- ZENG, L.; CEN, Y.; CHEN, J. Long non - coding RNA MALAT - 1 contributes to maintenance of stem cell - like phenotypes in breast cancer cells. **ONCOLOGY LETTERS**, v. 15, p. 2117–2122, 2018.
- ZHANG, Y.; WANG, T.; HUANG, H.; et al. Human MALAT-1 long non-coding RNA is overexpressed in cervical cancer metastasis and promotes cell proliferation , invasion and migration. **Jbuon**, v. 20, n. 238, p. 1497–1503, 2015.