

UNIVERSIDADE FEDERAL DO PARANÁ

MARCO ANTONIO CAMPANÁRIO SAMPAIO

INVESTIGAÇÃO DE POLIMORFISMOS DE NUCLEOTÍDEO ÚNICO EM RNAs
LONGOS NÃO-CODIFICANTES ASSOCIADOS À LEUCEMIA LINFOIDE AGUDA
INFANTIL

CURITIBA

2023

MARCO ANTONIO CAMPANÁRIO SAMPAIO

INVESTIGAÇÃO DE POLIMORFISMOS DE NUCLEOTÍDEO ÚNICO EM RNAS
LONGOS NÃO-CODIFICANTES ASSOCIADOS À LEUCEMIA LINFOIDE AGUDA
INFANTIL

Monografia apresentada como requisito parcial à
obtenção do título de Bacharel em Ciências
Biológicas, Curso de Ciências Biológicas, Setor de
Ciências Biológicas, Universidade Federal do
Paraná.

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

CURITIBA

2023

TERMO DE APROVAÇÃO PELA BANCA EXAMINADORA

A Lucie e Antonio, que me mostraram o lado bom da vida.

AGRADECIMENTOS

Agradeço à Universidade Federal do Paraná por ter me acolhido e por ser minha segunda casa ao longo dos 6 anos de graduação.

À minha orientadora, pela oportunidade de fazer parte do LabCHO e pela orientação gentil, humana e assertiva durante todo o processo.

Às minhas colegas de laboratório, pela ótima companhia e pelo suporte dado nos dias mais difíceis de trabalho.

Aos meus avós Lucie e Antonio por sempre estarem ao meu lado, protegendo-me.

À minha mãe Helena por ser meu exemplo de ser humano.

Ao meu irmão João por ser meu companheiro mais fiel na jornada da vida.

Aos meus tios René e Amanda pelo amor, acolhimento e parceria na cidade nova.

Ao Gaba, pela melhor companhia que eu poderia ter na faculdade e na vida.

À Marina, por todas as memórias compartilhadas desde o início da graduação.

Ao Matheus, por me lembrar todos os dias que o amor pode ser leve.

Ao Guiga, por acreditar em mim e reacender minha paixão pela dança.

À Ana, pela súbita conexão nessa cidade grande.

Ao elenco do Epifania Dance Project, pelas trocas diárias.

Aos meus amigos Gustavo, Leo, Jéssica, Érica, Gabriele e Alice, pelo companheirismo nos tempos bons e ruins.

Aos demais amigos e familiares que contribuíram para a conclusão desse trabalho.

À Resende, por ser meu porto seguro.

A mim e à minha criança interior.

RESUMO

Dentre os cânceres infantojuvenis, a leucemia é a neoplasia que mais acomete crianças e adolescentes no Brasil e no mundo. Ela pode ser classificada em quatro subtipos primários, dependendo do estado de maturação da célula cancerosa na medula óssea, sendo a leucemia linfóide aguda (LLA) o subtipo mais comum na idade pediátrica. Sabe-se que genes que codificam proteínas com funções de supressores tumorais e oncogenes têm papel fundamental na tumorigênese, mas recentes estudos têm ampliado o leque de moléculas que participam desse processo. Uma delas são os RNAs longos não codificantes (lncRNAs): transcritos de RNA maiores que 500 pares de base e sem informação para codificar proteínas. Eles fazem parte dos 98% do genoma não-codificante humano, e é exponencial a descoberta de novas variantes nessas moléculas associadas ao câncer. Devido a isso, e à carência de estudos de associação dessas moléculas com leucemia infantil na população brasileira, o presente estudo tem o objetivo de investigar polimorfismos associados à leucemia linfóide aguda (LLA). Para isso, foram selecionadas associações de polimorfismos de nucleotídeo único (SNPs) à LLA pediátrica já registradas em bases de dados de GWAS (estudos de associação genômica ampla), com base no filtro de *odds-ratio* > 1,2 e frequência do alelo menor > 0,05. Ao todo, 12 SNPs foram selecionados: rs7156960, rs630663, rs7738636, rs10018622, rs11155133, rs9290663, rs75777619, rs28665337, rs4617118, rs2069426, rs10170236 e rs11978267. Como bons candidatos para futuros estudos funcionais, duas regiões genômicas foram detalhadamente estudadas: *CDKN2B-AS1* (rs2069426) e *KCNMB2-AS1* (rs9290663). Para cada uma, foram desenhados *primers* alelo-específicos e, após a padronização das reações em cadeia da polimerase (PCRs), foi verificada a presença das variantes por sequenciamento de Sanger. As amostras caso consistiram em DNA extraído da fase leucocitária do sangue coletado de pacientes de 0-19 anos com leucemia, tratados no Hospital Oncopediátrico Erastinho. Já as amostras controle provieram de indivíduos saudáveis sem histórico familiar de câncer hereditário, armazenadas no Laboratório de Citogenética Humana e Oncogenética (LabCHO) da Universidade Federal do Paraná. Estudos como esse ampliam nossa compreensão sobre o papel dos lncRNAs na leucemia pediátrica, contribuindo para uma lacuna crucial na pesquisa oncológica.

Palavras-chave: LLA; lncRNA; SNP; GWAS; PCR alelo-específica.

ABSTRACT

Among childhood cancers, leukemia is the neoplasm that most affects children and teenagers in Brazil and worldwide. It can be classified into four primary subtypes, depending on the state of maturation of the cancerous cell in the bone marrow, with the acute lymphoblastic leukemia (ALL) being the most common subtype in the pediatric age group. It is known that genes encoding proteins with tumor suppressor functions and oncogenes play a crucial role in tumorigenesis. However, recent studies have expanded the range of molecules involved in this process. One example are the long non-coding RNAs (lncRNAs): RNA transcripts larger than 500 base pairs with no information for encoding proteins. They constitute part of the 98% of the non-coding human genome, and the discovery of new variants in these molecules associated with cancer is exponentially increasing. Due to this, and the lack of studies associating these molecules with childhood leukemia in the Brazilian population, the present study aims to investigate polymorphisms in lncRNAs associated with acute lymphoblastic leukemia (ALL). For this purpose, associations of single nucleotide polymorphisms (SNPs) with pediatric ALL already registered in genome-wide association studies (GWAS) databases were selected based on an odds ratio filter > 1.2 and minor allele frequency > 0.05 . In total, 12 SNPs were selected: rs7156960, rs630663, rs7738636, rs10018622, rs11155133, rs9290663, rs75777619, rs28665337, rs4617118, rs2069426, rs10170236, and rs11978267. As promising candidates for future functional studies, two genomic regions were thoroughly examined: CDKN2B-AS1 (rs2069426) and KCNMB2-AS1 (rs9290663). For each region, allele-specific primers were designed, and after standardizing polymerase chain reaction (PCR) reactions, the presence of variants was confirmed by Sanger sequencing. The case samples consisted of DNA extracted from the leukocyte phase of blood collected from patients aged 0-19 years with leukemia, treated at the Pediatric Oncology Hospital Erastinho. The control samples were derived from healthy individuals without a family history of hereditary cancer, stored in the Laboratory of Human Cytogenetics and Oncogenetics (LabCHO) at the Federal University of Paraná. Studies like this enhance our understanding of the role of lncRNAs in pediatric leukemia, contributing to a crucial gap in oncological research.

Key words: ALL; lncRNA; SNP; GWAS; allele-specific PCR

LISTA DE FIGURAS

Figura 1. Pureza do DNA extraído de amostras de sangue periférico de pacientes com LLA pediátrica.....	19
Figura 2. Região genômica do lncRNA ANRIL (CDKN2B-AS1).....	23
Figura 3. Região genômica do lncRNA KCNMB2-AS1.....	24
Figura 4. Gel de eletroforese da PCR da região flanqueadora do SNP rs2069426.	27
Figura 5. Cromatograma A do amplicon da região flanqueadora do SNP rs2069426.....	27
Figura 6. Cromatograma B do amplicon da região flanqueadora do SNP rs2069426.....	28
Figura 7. Alinhamento da sequência da região flanqueadora do SNP rs2069426 mapeada no genoma referência GRCh38.....	28
Figura 8. Gel de eletroforese da PCR da região flanqueadora do SNP rs9290663.	29
Figura 9. Cromatograma A do amplicon da região flanqueadora do SNP rs9290663.....	30
Figura 10. Cromatograma B do amplicon da região flanqueadora do SNP rs9290663.....	30
Figura 11. Alinhamento da sequência da região flanqueadora (226 pb) do SNP rs9290663 mapeada no genoma referência GRCh38.	31
Figura 12. Gel da padronização da PCR alelo-específica do SNP rs2069426 (G/T).	31
Figura 13. Gel de eletroforese da PCR alelo-específica do SNP rs2069426 (G/T) em amostras controle.	33
Figura 14. Gel de eletroforese da genotipagem do SNP rs2069426 (G/T) em amostras de DNA de LLA pediátrica.	33
Figura 15. Gel de eletroforese da genotipagem do SNP rs9092663 (A/T) em amostras de DNA de LLA pediátrica (amostras 01-10).	34
Figura 16. Gel de eletroforese da genotipagem do SNP rs9092663 (A/T) em amostras de DNA de LLA pediátrica (amostras 11-40).	34
Figura 17. Gel de eletroforese da genotipagem do SNP rs9092663 (A/T) em amostras de DNA de LLA pediátrica (amostras 41-57).	35

LISTA DE TABELAS

Tabela 1. SNPs associados a LLA pediátrica em lncRNAs.	20
Tabela 2. Parâmetros utilizados para desenho dos primers para PCR alelo-específica e PCR para sequenciamento da região flanqueadora do SNP rs2069426.	25
Tabela 3. Parâmetros utilizados para desenho dos primers para PCR alelo-específica e PCR para sequenciamento da região flanqueadora do SNP rs2069426.	26
Tabela 4. Condições para PCR da região flanqueadora do SNP rs2069426.	26
Tabela 5. Genótipos das amostras LLA01-07 a partir do sequenciamento de Sanger.	28
Tabela 6. Condições para PCR da região flanqueadora do SNP rs9290663.	29
Tabela 7. Genótipos das amostras LLA29-31 a partir do sequenciamento de Sanger.	31
Tabela 8. Condições para PCR alelo-específica do SNP rs2069426.	31
Tabela 9. Condições para PCR alelo-específica do SNP rs9290663.	32

LISTA DE QUADROS

Quadro 1. Quantificação e genotipagem das amostras de DNA extraído do sangue periférico de pacientes com LLA pediátrica (amostras LLA01-48).	42
Quadro 2. Quantificação e genotipagem das amostras de DNA extraído do sangue periférico de pacientes com LLA pediátrica (amostras LLA48-101).	43

SUMÁRIO

1. INTRODUÇÃO	9
2. REVISÃO DA LITERATURA	11
2.1. BIOLOGIA DO CÂNCER	11

2.2.	BIOLOGIA DAS LEUCEMIAS	12
2.3.	RNAs LONGOS NÃO CODIFICANTES	13
2.4.	ESTUDOS DE ASSOCIAÇÃO GENÔMICA AMPLA	14
3.	MATERIAL E MÉTODOS	15
3.1.	ANÁLISES IN SILICO	15
3.1.1.	EXTRAÇÃO DOS DADOS.....	15
3.1.2.	FILTRAGEM DOS DADOS	15
3.1.3.	SELEÇÃO DOS SNPS DE ESTUDO	15
3.1.4.	DESENHO DOS PRIMERS.....	16
3.2.	CASUÍSTICA	17
3.3.	EXPERIMENTOS <i>IN VITRO</i>	17
3.3.1.	EXTRAÇÃO DE DNA.....	17
3.3.2.	PCR PARA SEQUENCIAMENTO	18
3.3.3.	SEQUENCIAMENTO DE SANGER.....	18
3.3.4.	PCR PARA GENOTIPAGEM.....	18
4.	RESULTADOS.....	19
4.1.	QUANTIFICAÇÃO DAS AMOSTRAS DE SANGUE PERIFÉRICO	19
4.2.	SNPs CANDIDATOS PARA GENOTIPAGEM	20
4.6.	<i>PRIMERS</i> PARA A REGIÃO DE <i>rs9290663</i>	25
4.7.	SEQUENCIAMENTO REGIÃO FLANQUEADORA DE <i>rs2069426</i>	26
4.8.	SEQUENCIAMENTO REGIÃO FLANQUEADORA DE <i>rs9290663</i>	28
4.9.	PADRONIZAÇÃO DA PCR ALELO-ESPECÍFICA DE <i>rs2069426</i>	31
4.11.	GENOTIPAGEM DE <i>rs2069426</i>	32
4.12.	GENOTIPAGEM DE <i>rs9290663</i>	33
5.	DISCUSSÃO	35
6.	CONSIDERAÇÕES FINAIS.....	36
7.	REFERÊNCIAS BIBLIOGRÁFICAS	37
8.	APÊNDICE I	42

1. INTRODUÇÃO

O câncer é um conjunto heterogêneo de doenças complexas, constituindo a principal causa de mortes prematuras em todo o mundo (WILD, 2020). Especificamente em crianças e adolescentes (0-19 anos), a leucemia é a neoplasia com a maior taxa de ocorrência e mortalidade (INCA, 2020). As leucemias podem ser

categorizadas em quatro tipos principais, dependendo da maturação celular e da fase em que as células precursoras foram transformadas. Esses tipos incluem leucemia linfóide aguda (LLA), leucemia linfóide crônica (LLC), leucemia mieloide aguda (LMA) e leucemia mieloide crônica (LMC), conforme descrito pelo INCA em 2021. Dentre os subtipos primários, a LLA é o mais comum na idade pediátrica, abrangendo cerca de 80% dos casos.

Como em todo câncer, a proliferação desordenada de células é a característica central dessa patologia, o que a caracteriza como uma doença genética, embora nem sempre hereditária. Nesse sentido, uma vertente promissora na pesquisa oncológica envolve os RNAs longos não codificantes (lncRNAs), moléculas que, embora não codifiquem proteínas, desempenham um papel crucial na regulação da expressão gênica (YANG, LU & YUAN, 2014). Esses transcritos de RNA com tamanho maior que 500 pares de base estão envolvidos em importantes etapas do ciclo celular, e, assim como ocorre em regiões codificantes, polimorfismos em lncRNAs podem impactar no funcionamento celular (MATTICK *et al.*, 2023).

Sabendo disso, os estudos de associação genômica ampla (GWAS) emergem como ferramentas poderosas na identificação de variantes genéticas associadas a doenças – entre elas, a leucemia –, oferecendo uma visão abrangente do impacto dos polimorfismos em diferentes populações (WIJMENG & ZHERNAKOVA, 2018). A maioria dos GWAS publicados, disponibiliza os dados gerados, porém, foca suas análises e discussões em genes codificantes bem conhecidos, ficando novos lncRNAs pouco explorados. Adicionalmente, apesar do vasto volume de dados gerados por GWAS, a maioria dos dados depositados em bancos públicos é de ancestralidade europeia e, por consequência, a população brasileira não é bem representada (MILLS & RAHAL, 2020).

Portanto, o objetivo desse trabalho foi explorar dados de GWAS depositados em bancos de dados públicos em busca de polimorfismos mapeados em lncRNAs e associados à leucemia linfóide aguda pediátrica, a fim de selecionar algumas associações para validação em pacientes brasileiros. Estudos incipientes como este contribuem para o avanço do estudo da diversidade genética brasileira, bem como enriquecem o caminho em direção a abordagens mais precisas e individualizadas no prognóstico, diagnóstico e tratamento desse grupo complexo de doenças.

2. REVISÃO DA LITERATURA

2.1. BIOLOGIA DO CÂNCER

Câncer é um conjunto heterogêneo de mais de 100 doenças complexas e é a principal causa de mortes prematuras (antes dos 70 anos) na maior parte do mundo (WILD, 2020). Ele é caracterizado pela proliferação desordenada de células e engloba, em essência, doenças genéticas, porém não necessariamente hereditárias.

Hanahan (2022) e Weinberg (2000, 2011) compilaram as habilidades que uma célula deve apresentar para tornar-se cancerosa. Os *hallmarks* do câncer, como os autores denominam, são uma seleção de características comuns às células cancerosas, adquiridas durante sua vida pelo acúmulo de mutações. No total, 10 traços são listados como essenciais para a transformação celular. São eles: evasão de supressores de crescimento, imortalidade replicativa, resistência à morte celular, indução de angiogênese, sustentação de sinais proliferativos, inflamação promotora de tumor, instabilidade genômica, desregulação da energética celular, esquiva do sistema imune, e ativação da invasão e metástase.

Essas características confluem para o tema central da biologia do câncer: o ciclo celular desregulado das células cancerosas. Essa desordem é causada, entre outros fatores, por interferências nos circuitos de sinalização molecular. Isso tem diversas origens, e a oncogenética é um ramo da ciência que busca desvendá-las (WEINBERG, 2014).

Podem-se citar alguns caminhos pelos quais uma célula tem sua divisão celular desajustada. Mutações em genes supressores de tumor são alterações genéticas que contribuem para a carcinogênese, como o *TP53*, que codifica o fator de transcrição p53, importante proteína pró-apoptótica com atividade antitumoral pela via pRB/E2F1/p14-ARF/mdm2/p53 (PRIVES, 1998). Outra via conhecida da oncogênese é por meio de proto-oncogenes mutados (oncogenes), frequentes no genótipo de células tumorais. Um clássico exemplo é a família gênica *myc*, que codifica fatores de transcrição (c-, L-, N-myc) com atividade mitogênica constitutiva quando mutados, de forma a induzir a célula a entrar no ciclo celular – e a não sair – sem a necessidade de sinalização (AMATI, ALEVIZOPOULOS & VLACH, 1998).

Hoje, além dos caminhos mais populares da tumorigênese, como previamente citados, novas moléculas têm sido encontradas com papel importante no estabelecimento dos *hallmarks* do câncer. Uma delas são os RNAs longos não codificantes (lncRNAs), moléculas de RNA com mais de 500 pares de base, que não contêm informação para a tradução de proteínas funcionais completas (MATTICK *et al.*, 2023).

2.2. BIOLOGIA DAS LEUCEMIAS

No Brasil, estimam-se cerca de 625 mil novos casos de câncer para o triênio de 2020-2022 (INCA, 2020). Entre eles, a neoplasia que mais acomete crianças (0-15 anos) e adolescentes (15-19 anos) é a leucemia, e preveem-se cerca de 11 mil novos casos no país. Entre os cânceres infantojuvenis, a leucemia corresponde a 28% dos casos nos Estados Unidos (SIEGEL, MILLER & JEMAL, 2019), e entre todos os cânceres, ela representa 3,1% das mortes no mundo (SUNG *et al.*, 2021).

As leucemias podem ser divididas em quatro tipos primários, com base na maturação celular e em que fase tais células precursoras foram transformadas, sendo eles: leucemia linfóide aguda (LLA), leucemia linfóide crônica (LLC), leucemia mieloide aguda (LMA) e leucemia mieloide crônica (LMC) (INCA, 2021).

Leucemias mielóides têm origem na transformação de mieloblastos – células precursoras de megacariócitos, eritrócitos e leucócitos –, enquanto leucemias linfóides surgem a partir da transformação de linfoblastos – células precursoras de linfócitos. A classificação crônica/aguda depende da fase em que o câncer surgiu nos blastos e da capacidade de maturação deles. Se surgiu em fase imatura, impedindo a maturação das células, a leucemia é aguda, e se surgiu em fase imatura, porém o blasto tem capacidade de maturação, a leucemia é crônica. (AMERICAN CANCER SOCIETY, 2018).

Existem diferentes subtipos de leucemia, inclusive aqueles que não se enquadram nos tipos primários, como a leucemia mielomonocítica crônica e a síndrome mielodisplásica. Os subtipos são categorizados também a partir das anomalias citogenéticas recorrentes, como aneuploidias e translocações (ARBER *et al.*, 2016). A 5ª edição do *World Health Organization Classification Of Haematolymphoid Tumours* já foi revisada por alguns grupos de pesquisadores

(ARBER *et al.*, 2022; ALAGGIO *et al.*, 2022; KHOURY *et al.*, 2022), e a própria comunidade científica concorda que a classificação dos subtipos leucêmicos é uma “confusão babilônica” (HUBER *et al.*, 2023).

2.3. RNAs LONGOS NÃO CODIFICANTES

Em sua maioria, os lncRNAs são sintetizados pela RNA polimerase II (RNA Pol II) e passam por um processo de maturação assim como os RNAs mensageiros (mRNA – *messenger RNA*), ou seja, possuem cap 5', cauda poli-A e sofrem *splicing*. Em comparação com mRNAs, os lncRNAs têm maior comprimento, porém possuem menos éxons e menor conservação evolutiva. Apesar de alguns codificarem pequenos peptídeos, a grande maioria não possui ORFs (*open reading frames* – quadros abertos de leitura), porém, essas moléculas têm papel chave na regulação da expressão gênica (STATELLO *et al.*, 2021).

Os lncRNAs podem estar localizados no núcleo, no citoplasma, na mitocôndria, em outras organelas ou até associados a ribossomos, dependendo da característica estrutural da molécula. Existem mecanismos de retenção de lncRNAs no núcleo, como por exemplo o bloqueio da exportação de moléculas com um a poucos éxons para o citoplasma (via NXF1) ou a interação do lncRNA recém-sintetizado com proteínas da cromatina. Dependendo da sua localização celular, esses RNAs atuam em diferentes níveis da expressão gênica (BRIDGES, DAULAGALA & KOURTIDIS, 2021).

Esses RNAs podem ser categorizados com base em sua forma (circRNAs – lncRNAs circulares), história evolutiva (pseudogenes), localização genômica (lincRNAs – lncRNAs intergênicos; lncRNAs antissenso) e em outros fatores como revisado por Laurent, Wahlestedt e Kapranov (2015). Os lncRNAs antissenso são aqueles antiparalelos a um ou mais genes e podem ser capazes de modulá-los em *cis* ou em *trans* por diferentes mecanismos, como esponja de microRNAs e recrutador de marcas epigenéticas, por exemplo (VILLEGAS & ZAPHIROPOULOS, 2015).

Assim como em regiões codificantes, variantes genéticas de lncRNA podem ter relação com o fenótipo de doenças complexas, visto que a maioria das associações encontradas em estudos de associação genômica ampla (GWAS) são mapeadas em regiões não codificantes (ZHANG & LUPSKI, 2015).

2.4. ESTUDOS DE ASSOCIAÇÃO GENÔMICA AMPLA

GWAS (*Genome-Wide Association Study*) é um modelo de estudo genético livre de hipótese que investiga a associação entre variações genéticas comuns em todo o genoma e a presença de características ou doenças específicas em uma população. Os GWAS são uma ferramenta valiosa na genômica para compreender a base genética de uma variedade de fenótipos e contribui para avanços na medicina personalizada e na compreensão das bases genéticas de várias condições de saúde (IKEGAWA, 2012).

A era das ciências “ômicas” trouxe consigo a explosão de GWAS, amplificando a capacidade de pesquisa genética com coortes de grande porte, como populações de mais de 100.000 indivíduos, mas também carregando importantes desafios, como questões éticas e interpretativas (WIJMENGA & ZHERNAKOVA, 2018).

Apesar da exponencial produção de dados genéticos populacionais, ainda há assimetria e hegemonia entre as ancestralidades representadas nos bancos de dados (POPEJOY & FULLERTON, 2016). Em 2017, 88% de todos os dados gerados por GWAS, depositados tanto no US *National Human Genome Research Institute* (NHGRI) quanto no *European Bioinformatics Institute* (EBI), eram de ancestralidade europeia (MILLS & RAHAL, 2019). Mills e Rahal (2020) tentam apontar tais discrepâncias através da ferramenta *GWAS Diversity Monitor*, um painel em tempo-real de dados de GWAS depositados nos principais bancos de dados, com filtros de doenças e ancestralidades.

Além disso, com já dito, a maior parte dos dados gerados por GWAS são mapeados em regiões não codificantes do genoma (ZHANG & LUPSKI, 2015). Como GWAS geram associações estatísticas apenas, sem explicação funcional, cabe à comunidade científica realizar estudos experimentais para validar a função biológica das associações. Essa ponte entre associação e função é escassa para lncRNAs (GALLAGHER & CHEN-PLOTKIN, 2018), o que reflete uma área de pesquisa promissora, visto que essas regiões genômicas são *hotspots* de GWAS e diversos mecanismos moleculares de tumorigênese são elucidados através deles (WANG *et al*, 2023, *pre-print*).

3. MATERIAL E MÉTODOS

3.1. ANÁLISES IN SILICO

3.1.1. EXTRAÇÃO DOS DADOS

Os dados de associação genética foram extraídos da plataforma GWAS Catalog (BUNIELLO *et al.*, 2019). Cada *dataset* (conjunto de dados) da plataforma é identificado com um EFO ID (Experimental Factor Ontology Identifier), cuja finalidade é categorizar os estudos depositados no banco de dados de acordo com suas variáveis experimentais. Pelo sistema de busca do catálogo, os termos “ALL”, “Acute Lymphoid Leukemia”, “Acute Lymphoblastic Leukemia”, “AML”, “Acute Myeloid Leukemia” e “Acute Myeloblastic Leukemia” foram pesquisados e somente os resultados dentro da categoria *traits* (fenótipos) foram considerados.

3.1.2. FILTRAGEM DOS DADOS

Os *datasets* selecionados para pesquisa foram compilados, e os polimorfismos associados à leucemia mieloide aguda (LMA), crônica (LMC) e à leucemia linfóide crônica (LLC) foram excluídos da amostra de estudo, a fim de restarem apenas variantes estatisticamente associadas à leucemia linfóide aguda (LLA). Além disso, foram excluídas as variantes associadas a fenótipos secundários em pacientes com LLA, como resistência ou hipersensibilidade a quimioterápicos (exemplo: variantes associadas à hipersensibilidade à asparaginase em pacientes com LLA).

O *software BioMart* da plataforma *Ensembl* (CUNNINGHAM *et al.*, 2022) foi usado para recuperar informações sobre a entidade genômica onde localiza-se cada polimorfismo da amostra. Dessa forma, foi possível identificar e selecionar apenas os SNPs localizados em lncRNAs a partir do identificador gênico (*Gene stable ID*; exemplo: ENSG00000229140).

3.1.3. SELEÇÃO DOS SNPS DE ESTUDO

Os estudos que identificaram associações genéticas de variantes de lncRNAs com LLA foram lidos e analisados quanto à idade da coorte, subtipo de LLA, MAF (*minor frequency allele*) do SNP e valor de OR (*odds ratio*) da associação estatística. Os SNPs com MAF maior que 0,05, OR maior que 1,2 ou menor que 0,8, e

encontrados em estudos de coorte com idade menor que 19 anos (crianças, adolescentes e jovens adultos) foram selecionados para o desenho de *primers*.

Para a análise dos SNPs, foram usadas as plataformas *Ensembl* (CUNNINGHAM, 2022) e *ABraOM* (NASLAVSKY *et al.*, 2022) para dados populacionais, *dbSNP-NCBI* (*Database of Single Nucleotide Polymorphisms*) (SHERRY, WARD & SIROTKIN, 1999) para dados genômicos e o *UCSC Genome Browser* (KENT *et al.*, 2002) para visualização do contexto genômico.

3.1.4. DESENHO DOS PRIMERS

Duas estratégias foram adotadas como tentativa para as PCRs alelo-específicas: PCR tradicional e qPCR *multiplex*. Para PCR tradicional, foi feito o desenho de *primers forward* alelo-específicos e o uso de um mesmo *primer reverse* para reações separadas (ou vice-versa). Já para a PCR *multiplex*, um *primer forward* alelo-específico e um *reverse* alelo-específico foram desenhados para uma mesma reação de PCR em tempo real (qPCR), de forma que o *primer forward* alelo-específico amplifique um fragmento de um certo tamanho, e o *reverse* alelo-específico amplifique um outro fragmento de tamanho diferente, para que ao final de cada reação, possa ser feita a análise por curva de *melting* e a genotipagem por qPCR *multiplex*.

Para ambos os casos, a sequência flanqueadora de cada SNP (aproximadamente 1000 pb) foi baixada em *fasta* na plataforma *dbSNP* do NCBI (SHERRY, WARD & SIROTKIN, 1999). Para o desenho dos *primers* alelo específicos (tanto *forward* quanto *reverse*), a sequência flanqueadora de cada SNP foi inserida na ferramenta *Primer-Blast* do NCBI (YE, 2012) como PCR *template* e uma sequência de 19-25 pb contendo o alelo específico terminal foi inserida como “*own primer*”. No caso dos *primers reverse* alelo-específicos, a sequência foi transformada em sua reversa complementar para inserção na ferramenta.

Os 10 *primers* encontrados para cada *forward* e *reverse* alelo-específicos foram analisados quanto à temperatura de *melting* (T_m) ($\Delta T_m < 5,0$), autocomplementaridade (*self-complementarity* $< 2,0$), autocomplementaridade 3' (*self-3'-complementarity* $< 2,0$) e conteúdo GC (%GC $\sim 55\%$). Os *primers* que atingiram os valores de cada parâmetro foram analisados na ferramenta *OligoAnalyzer* (IDT™ *Integrated DNA Technologies*) sob os parâmetros de qPCR.

Nessa ferramenta, foi analisada a estrutura tridimensional da molécula para verificar a possibilidade de formação de *hairpins* ($\Delta G < -6$; $\Delta G > 6$), *self-dimers* (~ 0) e *hetero-dimers* (~ 0) entre as moléculas de cada PCR. Os *primers* que atingiram os valores estabelecidos foram pedidos para síntese.

Para confirmar os alelos presentes na população brasileira paranaense, foram desenhados *primers* flanqueadores de cada região dos SNPs candidatos, com o SNP localizado no centro. Os parâmetros para desenho dos *primers forward* e *reverse* foram os mesmos citados anteriormente, e os *amplicons* foram sequenciados pelo método de Sanger.

3.2. CASUÍSTICA

Para a genotipagem dos SNPs selecionados, serão analisadas as amostras de pacientes com leucemia provenientes do Hospital Oncopediátrico Erastinho e pacientes controles provenientes do banco de amostras negativas de COVID do Laboratório de Citogenética Humana e Oncogenética (LabCHO) da Universidade Federal do Paraná (UFPR) (CAAE: 43945121.4.0000.0102). As amostras de câncer são coletadas sob aprovação do Comitê de Ética em Pesquisa da Liga Paranaense de Combate ao Câncer - Hospital Erasto Gaertner, sob o número CAAE: 08809419.0.0000.0098 em 10/04/2019. Todos os pacientes assinaram o Termo de Consentimento Livre e Esclarecido (TCLE) e o Termo de Assentimento Livre e Esclarecido (TALE) para pacientes menores de idade.

3.3. EXPERIMENTOS *IN VITRO*

3.3.1. EXTRAÇÃO DE DNA

As amostras de leucemia estão armazenadas a $-4\text{ }^{\circ}\text{C}$ no Laboratório de Citogenética Humana e Oncogenética (LabCHO). Para seu processamento, elas foram descongeladas, centrifugadas, e 200 μL do sangue total é recuperado para extração de DNA. O DNA foi extraído seguindo o protocolo 01 (sangue total humano) do Biopur™ Kit Mini Spin Plus. Ao final da extração, o *pellet* na coluna foi eluído em 50 μL da solução de eluição. Cada amostra é quantificada em espectrofotômetro NanoDrop® 2000 (Thermo Scientific).

3.3.2. PCR PARA SEQUENCIAMENTO

A região flanqueadora dos SNPs candidatos foi amplificada a partir dos *primers* desenhados. As PCRs foram montadas a partir de 2 µl do DNA *template* (20 ng/µL, 0,5 µL da solução contendo os *primers forward* e *reverse* (100 mmol/100 µL), 17,5 µL de água ultrapura e 10µL de Master Mix para PCR convencional (1x) desenvolvido pelo IBMP, ICC/FioCruz.

3.3.3. SEQUENCIAMENTO DE SANGER

Para o sequenciamento dos *amplicons* das regiões flanqueadoras dos SNPs, primeiramente eles foram quantificados em espectrofotômetro NanoDrop® 2000 (Thermo Scientific). Depois, as amostras foram tratadas com exonuclease (exo1) e fosfatase alcalina (SAP). Para cada 9 µL de produto de PCR, foram adicionados 0,33 µL de exo1, 0,33 µL de SAP e 0,34 µL de água ultrapura. Essa solução foi colocada em termociclador a 37°C por 1 hora, e depois a 85°C por 15 minutos.

Após essa etapa, seguiu-se o protocolo padrão de marcação fluorescente pelo *BigDye™ Terminator*. Para cada 4 µL de produto de PCR tratado com exo/SAP, foram adicionados 2,0 µL de tampão *BigDye™*, 2,6 µL de água ultrapura e 0,7 µL de solução de *primer* (*forward* e *reverse*) a 10 pmol/µL. Essa solução foi colocada em um termociclador seguindo 40 ciclos de 15 segundos a 96°C para desnaturação, 15 segundos a 64°C para anelamento e 4 minutos a 60°C para extensão.

A última fase consistiu na purificação do DNA marcado por precipitação, etapa em que EDTA (125 mM) e etanol (absoluto e 70%) foram utilizados para sessões de centrifugação e *spins* invertidos, a fim de precipitar o produto de PCR marcado. Após a evaporação completa do etanol, as amostras são vedadas e enviadas ao laboratório sequenciador.

3.3.4. PCR PARA GENOTIPAGEM

A região flanqueadora do SNP rs2069426 (491 pb) e do SNP rs9290663 (226 pb) foi amplificada a partir dos *primers* desenhados. As PCRs foram montadas a partir de 40 ng do DNA *template*, 0,5 µL da solução contendo os *primers forward* e *reverse*

(100 mmol/100 μ L), 17,5 μ L de água ultrapura e 10 μ L de Master Mix para PCR convencional (1x) desenvolvido pelo IBMP, ICC/FioCruz.

O produto da PCR foi genotipado por meio do padrão eletroforético dos fragmentos amplificados em gel de agarose 1,5% suplementado com *Gel Red* (Biotium). O gel de eletroforese foi interpretado a partir da presença ou ausência dos produtos de PCR no gel.

4. RESULTADOS

4.1. QUANTIFICAÇÃO DAS AMOSTRAS DE SANGUE PERIFÉRICO

A quantificação das amostras de DNA extraído do sangue periférico dos pacientes com LLA está apresentada na tabela I, com o código das amostras, concentração em ng/ μ L e razão 260/280 (*vide* Apêndice I). Ao todo, 82 amostras foram coletadas e quantificadas, mas 13 delas não apresentaram conteúdo de ácido nucleico nas leituras, sendo necessário extraí-las novamente em estudos futuros. A média de pureza das amostras foi de 1,78, com desvio-padrão de 0,19 (Figura 1). Para a produção de alíquotas das amostras para uso nos experimentos, as soluções foram diluídas à concentração de 20 ng/ μ L (somente aquelas com concentração maior que 20 ng/ μ L).

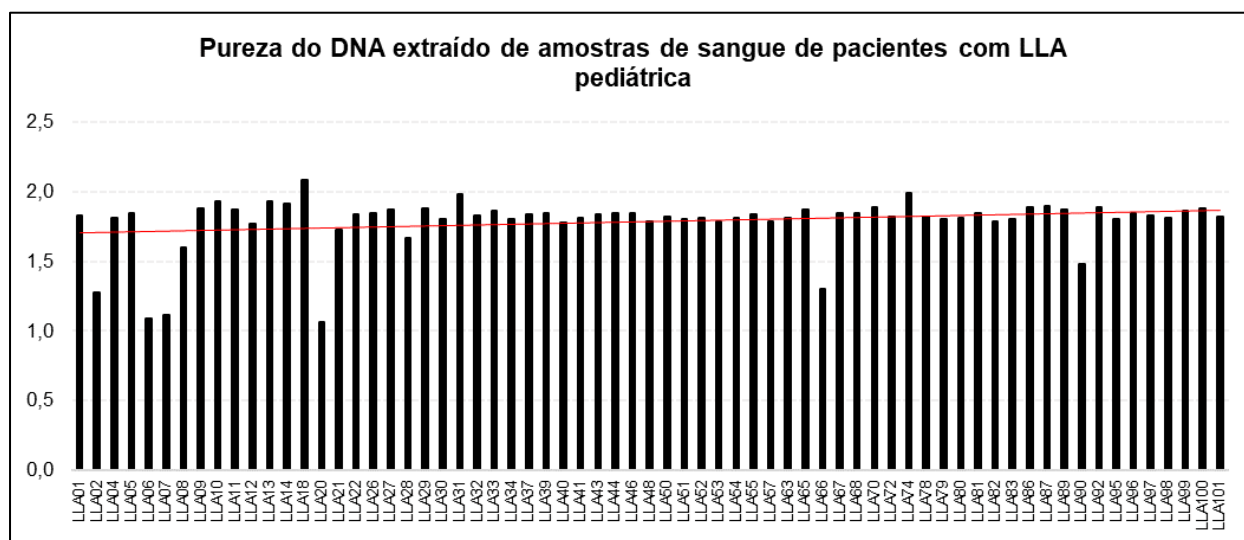


Figura 1. Pureza do DNA extraído de amostras de sangue de pacientes com LLA pediátrica. Linha vermelha: média da razão 260/280.

4.2. SNPs CANDIDATOS PARA GENOTIPAGEM

Ao todo, dados de 58 GWAS foram filtrados em busca de SNPs mapeados em lncRNAs e associados à LLA pediátrica. Desses 58 GWAS, 445 associações foram filtradas, sendo 257 delas associadas à LLA e 158 associadas à LLC. Das 257, apenas 137 eram associadas à LLA pediátrica. Dentre as 137 associações à LLA pediátrica, 33 SNPs foram encontrados mapeados em 32 lncRNAs diferentes. Após os filtros de *odds-ratio* ($> 1,2$) e MAF ($> 0,05$), restaram 12 SNPs mapeados em 10 lncRNAs diferentes (Tabela 1).

Nesse sentido, com base nos critérios mencionados, foram selecionados 12 SNPs de lncRNAs associados à LLA pediátrica: rs7156960, rs630663, rs7738636 (ELLINGHAUS, 2011), rs10018622 (ARCHER, 2017), rs11155133, rs9290663 (TREVINO, 2009), rs75777619, rs28665337 (VIJAYAKRISHNAN, 2019), rs4617118, rs2069426 (WIEMELS, 2018), rs10170236 (YANG, 2012) e rs11978267 (ELLINGHAUS, 2011; TREVINO, 2009; WIEMELS, 2018). O SNP rs11978267 é citado em três estudos diferentes (Tabela 1).

Tabela 1. SNPs associados a LLA pediátrica em lncRNAs.

SNP	lncRNA	OR	CI (95%)	MAF	Alleles	GWAS
rs2069426	<i>CDKN2B-AS1</i>	1.35	[1.26 - 1.70]	0.06 (G)	G T	Wiemels <i>et al.</i> , 2018
rs75777619	<i>CCDC26</i>	1.26	[1.17 - 1.36]	0.1 (A)	A G	Vijayakrishnan <i>et al.</i> , 2019
rs28665337	<i>CCDC26</i>	1.34	[1.21 - 1.47]	0.1 (C)	C A T	Vijayakrishnan <i>et al.</i> , 2019
rs4617118	<i>CCDC26</i>	1.28	[1.19 - 1.37]	0.19 (A)	A G	Wiemels <i>et al.</i> , 2018
		1.69	[1.40 - 1.90]	0.23 (G)	A G	Trevino <i>et al.</i> , 2009
rs9290663	<i>KCNMB2-AS1</i>	1.58	[1.20 - 1.90]	0.25 (A)	A T G	Trevino <i>et al.</i> , 2009
rs11978267	<i>AC124014.1</i>	1.43	[1.28 - 1.59]	0.23 (G)	A G	Ellinghaus <i>et al.</i> , 2011
		1.43	[NR]	0.23 (G)	A G	Wiemels <i>et al.</i> , 2018
rs7156960	<i>AC016526.2</i>	1.2	[1.12 - 1.33]	0.44 (G)	G C	Ellinghaus <i>et al.</i> , 2011
rs630662	<i>AC025508.1</i>	[NR]	[NR]	0.2 (G)	G A T	Ellinghaus <i>et al.</i> , 2011
rs10018622	<i>AC079921.1</i>	1.9	[1.45 - 2.63]	0.49 (G)	G A C	Archer <i>et al.</i> , 2017
rs11155133	<i>AL035446.1</i>	3.62	[2.10 - 6.00]	0.13 (A)	A G	Trevino <i>et al.</i> , 2009
rs7738636	<i>AL355612.1</i>	1.2	[1.15 - 1.41]	0.28 (A)	A C	Ellinghaus <i>et al.</i> , 2011
rs10170236	<i>MMADHC-DT</i>	1.45	[1.24 - 1.69]	0.33 (A)	A G	Yang <i>et al.</i> , 2012

A maior parte dos SNPs selecionados estão mapeados em lncRNAs de função desconhecida (AC124014.1, AC016526.1, AC025508.1, AC079921.1, AL035446.1, AL355612.1, MMADHC-DT). Os demais SNPs estão mapeados em lncRNAs com registros na literatura de estudos funcionais com linhagens de diversos cânceres, como câncer de pâncreas (PENG *et al.*, 2016), de tireoide (MA *et al.*, 2021) e

glioblastoma (WEI *et al.* 2014) para o lncRNA *CCDC26*, que apresentou 3 SNPs diferentes associados à LLA pediátrica em sua sequência.

Pela importância biológica já registrada na literatura, neste trabalho, serão apresentados os resultados referentes aos SNPs rs2069426 (*CDKN2B-AS1* / *ANRIL*) e rs9290663 (*KCNMB2-AS1*), pois os mesmos foram selecionados para análise na população de Curitiba. O contexto genômico de cada uma dessas regiões está descrito nas sessões seguintes.

4.3. CONTEXTO GENÔMICO DE *rs2069426*

A variante rs2069426 está localizada no lncRNA antissenso *CDKN2B-AS1*, também conhecido por *ANRIL* ou *p15AS* (Figura 2). Ele está localizado no braço pequeno do cromossomo 9 humano (9p21.3) e possui 133.964 pb. Ao longo de sua sequência, existem sequências Alu, SINEs (*short interspersed nuclear elemento*) e LINEs (*long interspersed nuclear elements*). Além disso, essa é a região do cluster gênico *p16/p15/p14 (INKb-ARF-INKa)*, a montante do *ANRIL*. Inclusive, o gene da proteína p14-ARF compartilha um promotor bidirecional com o lncRNA, que é reconhecido por E2F1 (ENSEMBL, 2021).

Os transcritos de *ANRIL* são RNAs funcionais – incluindo variantes de *splicing* circulares – de cerca de 3800 pb que interagem com PRC1 e PRC2 na cromatina. O recrutamento dessas proteínas pelo lncRNA favorecem o silenciamento de *INKb-ARF-INKa* por marcas epigenéticas repressivas. PRC2 induz a tripla metilação da lisina 27 na histona H3 (H3K27met3) através da atividade enzimática de EZH2. Por sua vez, PRC1 reconhece essa marca e ativa a monoubiquitinação de histonas H2A (YAP *et al.*, 2010).

Nesse sentido, o *cluster INKb-ARF-INKa* tem sua expressão reprimida pela ação do *ANRIL* em condições celulares normais. Isso porque INKb (p16), INKa (p14) e p14 (ARF) atuam como supressores tumorais, ou seja, seu mecanismo de ação deve ser desencadeado apenas quando necessário. Esse mecanismo se dá pela interação de p15 e p16, quando ativas, com as ciclinas CDK4 e CDK63, prevenindo a fosforilação de RB. Isso é um sinal para a fase G1 do ciclo celular ser abortada. Concomitantemente, ainda, p14, quando ativa, sequestra mdm2 no nucléolo e induz apoptose mediada por p53.

Sabe-se que esse lncRNA antissenso tem sua expressão desregulada em uma série de doenças humanas, incluindo câncer (SONG *et al.*, 2020). Por exemplo, em células de câncer de mama triplo-negativo, esse transcrito foi encontrado com expressão hiper-regulada (IRANPOUR *et al.*, 2015). Em leucemia linfóide aguda de células T (LLA-T), mais de 70% dos pacientes apresentam deleção do *cluster INKb-ARF-INKa*, incluso *ANRIL*. Além disso, estudos de associação genômica ampla (GWAS – *genome-wide association study*) têm achado variantes genéticas no *ANRIL* associadas à suscetibilidade a diferentes tipos de leucemia (WIEMELS *et al.*, 2018).

A figura 2 mostra diversas alças de interação entre elementos regulatórios da região genômica do *ANRIL*, incluindo a região específica do rs2069426. Wang *et al.* (2020) identificaram que o microRNA *miR-34a* é um alvo *downstream* do *ANRIL*. Os autores mostraram que o *ANRIL* é regulado positivamente em células de LMA, sequestrando o *miR-34a* e aumentando a capacidade de proliferação, migração e invasão das células. Além disso, também é sabido de um mecanismo similar em células de LLA de células T. Por meio de experimentos com linhagens celulares de LLA (Jurkat E6-1 e CCRF-CEM), foi mostrado que o silenciamento do *ANRIL* atenua a carcinogênese de LLA-T pediátrica. Isso porque o lncRNA regula a expressão de *TRAF5* também através do sequestro de outro microRNA (*miR-335-3p*). O gene *TRAF5* está associado à tumorigênese de LLA-T por atuar nas vias de sinalização do NF- κ B e MAPK (CHEN *et al.*, 2023).

Além disso, o *ANRIL* é considerado um *hotspot* de GWAS inesperado, ou seja, essa região aparece associação a diversos fenótipos complexos nos estudos ao redor do mundo, mas cabe-se agora entender os motivos funcionais por trás desse fato (PASMANT *et al.*, 2011).

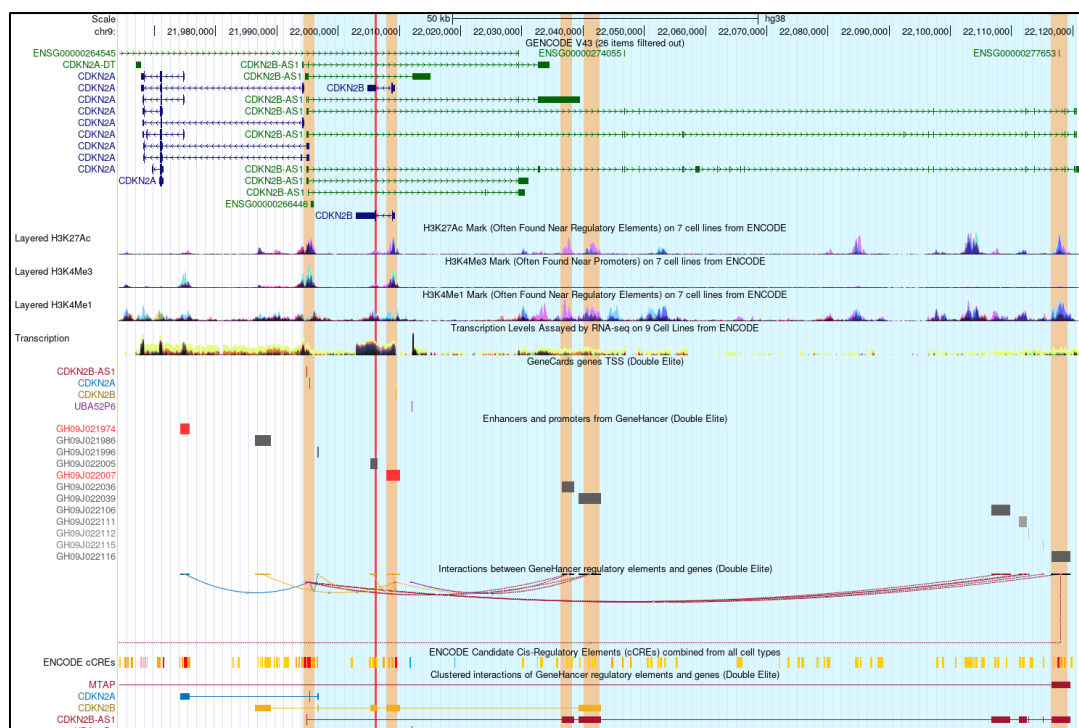


Figura 2. Região genômica do lncRNA ANRIL (CDKN2B-AS1). Região azul: ANRIL. Região laranja: Picos de marcas epigenéticas (H3K27Ac, H3K4Me3; H3K4Me1). Região vermelha: rs2069426.

4.4. CONTEXTO GENÔMICO DE *rs9290663*

A variante *rs9290663* está localizada no lncRNA *KCNMB2-AS1*, no cromossomo 3, na posição 178.711.787 da fita *forward* (Figura 3). Esse lncRNA é estudado em diversos outros tipos de câncer, como carcinoma de pulmão de células não pequenas (NSCLC) (YANG & WANG, 2020), de bexiga (ZHU *et al.*, 2021; CHEN *et al.*, 2021), cervical (ZHANG *et al.*, 2020) e de endométrio (TANG *et al.*, 2019). Nesses estudos, observou-se a regulação positiva de processos moleculares carcinogênicos e o lncRNA.

Especificamente em relação ao polimorfismo, o *rs9290663* foi avaliado por Cao *et al.* (2016) em uma coorte da China. Nesse trabalho, foi mostrado que o alelo T dessa variante pode ter relação com o aumento da susceptibilidade à leucemia mieloide aguda, a partir do estudo de seu impacto funcional no íntron do lncRNA. Por ser antissenso do gene *KCNMB2*, o íntron do lncRNA está sobreposto à região codificante da subunidade $\beta 2$ de um canal de potássio cálcio-dependente. Nesse sentido, é possível que essas duas moléculas interajam, com base no modelo genético usado pelos autores.

De fato, é destacável a região genômica *KCNMB2/KCNMB2-AS1* em GWAS, como mostrado na revisão de Marrero e Lamba (2023). O gene *KCNMB2* codifica um canal proteico essencial para controlar o fluxo de íons de potássio entre membranas, desempenhando funções em processos fisiológicos diversos, como a excitabilidade neuronal, a contração suave dos músculos e a secreção hormonal. O porquê de essa região aparecer frequentemente associada à leucemia pediátrica ainda é uma lacuna na literatura.

4.5. PRIMERS PARA A REGIÃO DE *rs2069426*

Foram encontrados *primers* alelo-específicos adequados (*vide* Metodologia) para o SNP *rs2069426*. Ele é bialélico (G/T, *Ensembl*) e tem MAF de 0,06 (T) com base no conjunto de dados da fase 3 do *1000 Genomes Project* (FAIRLEY *et al.*, 2020) (Tabela 2). Porém, no trabalho de Wiemels *et al.* (2018), o alelo A/C dessa variante é que está associado à LLA ($p = 3,7 \times 10^{-9}$). Em vista disso, o sequenciamento da região flangeadora do SNP foi realizado para confirmar os alelos presentes nas amostras do estudo.

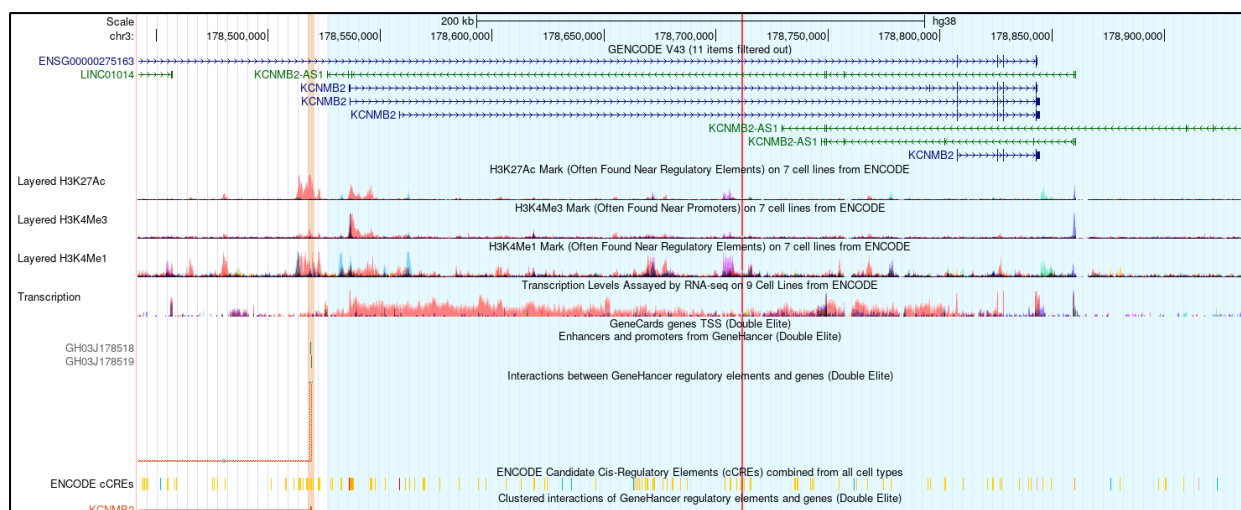


Figura 3. Região genômica do lncRNA *KCNMB2-AS1*. Região azul: *KCNMB2-AS1*. Região laranja: Picos de marcas epigenéticas (H3K27Ac, H3K4Me3; H3K4Me1). Região vermelha: *rs9290663*.

No *Ensembl*, o único conjunto de dados que apresentou o alelo A/C genotipado, foi o *genomAD exomes r2.1.1* da *Genome Aggregation Database* (*genomAD*), em que dentre 113.642 indivíduos genotipados (227.284 alelos), apenas

1 alelo A e 1 alelo C foram encontrados, ambos em uma população do Leste Asiático, com frequência de 0,0000044 cada.

No estudo de Wiemels *et al.* (2018), os autores realizaram um GWAS de caso-controle estratificado em 3 grupos étnicos (latinos, brancos não-latinos e afro-americanos). A coorte de estudo foi composta de 3263 casos do *California Childhood Cancer Records Linkage Project* (CCRLP) – 1949 latinos, 1184 brancos não-latinos e 130 afro-americanos – e 15.977 controles, sendo 3.506 do CCRLP – 2.120 latinos, 1.241 brancos não-latinos e 145 afro-americanos – e 12.421 do Kaiser Genetic Epidemiology Research on Aging (GERA) – 6.464 latinos, 2.310 brancos não-latinos e 3697 afro-americanos. Os resultados foram replicados em dois *datasets*: uma replicação em uma coorte de brancos não-latinos do *Children's Oncology Group* (COG, protocolos 9904 e 9905) com 959 casos fundida com uma coorte também de brancos não-latinos do *Welcome Trust Case Control Consortium* (WTCCC) com 2.624 controles, e outra replicação em uma coorte de latinos do *California Childhood Leukemia Study* (CCLS) com 530 casos e 511 controles.

No material suplementar do estudo de Wiemels *et al.* (2018), são mostrados os dados da coorte do CCRLP relativos aos SNPs mais relatados na literatura prévia, e entre eles está o alelo A/C do rs2069426. Sobre ele, são relatados o p-valor, o *odds ratio* (OD) com intervalo de confiança de 95% e o RAF (*risk allele frequency*). Para o alelo A/C, o RAF dos latinos do CCRLP é de 0,045, o dos brancos não-latinos é de 0,081 e o dos afro-americanos é de 0,099, todos maiores que o valor mostrado no Ensembl (0,0000044).

Tabela 2. Parâmetros utilizados para desenho dos primers para PCR alelo-específica (fragmento de 189 pb) e PCR para sequenciamento da região flanqueadora do SNP (491 pb) rs2069426.

rs2069426 (CDKN2B-AS1)	Forward (G)	Forward (T)	Forward (flanker)	Reverse
Primer sequence	CAGAGAGCAGAGTGGTCAGAG	CAGAGAGCAGAGTGGTCAGAT	AAAATAAAGTCGTTGTGGGCGG	TGAAGGTGGGGTGGGAAAGT
GC (%)	56.5	52.2	45.5	55
Tm (°C)	64.9	64.8	63.8	65.6
Hairpin (ΔG)	-2.27	-2.27	-1.42	3.49
Self-dimer 3' (ΔG)	-3.14	-3.14	-3.61	0

4.6. PRIMERS PARA A REGIÃO DE rs9290663

Foram encontrados *primers* alelo-específicos adequados (*vide* Metodologia) para o SNP rs9290663 (Tabela 3). Existem 3 alelos do SNP (A/G/T, Ensembl) e ele

tem MAF de 0,25 (T) com base no conjunto de dados da fase 3 do 1000 Genomes Project (FAIRLEY *et al.*, 2020). Em vista disso, o sequenciamento da região flanqueadora do SNP também foi feito para verificar os alelos presentes nas amostras de estudo.

Tabela 3. Parâmetros utilizados para desenho dos primers para PCR alelo-específica (fragmento de 154 pb) e PCR para sequenciamento da região flanqueadora do SNP (226 pb) rs2069426.

rs9290663 (KCNMB2-AS1)	Forward	Reverse (A)	Reverse (T)	Reverse (flanker)
Primer sequence	GCCGTTGTCACGTGTGAGTAG	CGGAGCTACCACTACTA ^A	CGGAGCTACCACTACTA ^T	ATGGCACGTAGGTGGTTAGG
GC (%)	55.0	50.0	50.0	55
Tm (°C)	62.3	60.4	60.2	63.6
Hairpin (ΔG)	-0.14	-1.28	-1.28	-2.03
Self-dimer 3' (ΔG)	-4.87	-6.34	-6.34	-6.3

4.7. SEQUENCIAMENTO REGIÃO FLANQUEADORA DE rs2069426

O gel de eletroforese do *amplicon* da região flanqueadora (491 pb) do SNP rs2069426 mostrou bandas do tamanho esperado em relação ao que foi desenhado *in silico* (Figura 4). As condições da reação estão mostradas na tabela 4. Os cromatogramas do sequenciamento desses *amplicons* revelaram dois picos de fluorescência na posição do SNP para algumas amostras testadas, indicando genótipo heterozigoto (G/T) (Figuras 5-6). Ao todo, 7 amostras de LLA foram sequenciadas para esse teste (Figura 7). Com isso, a incongruência dos dados do GWAS e do Ensembl foi resolvida, mostrando que, pelo menos nos indivíduos do estudo, os alelos presentes são G/T (Tabela 5).

Tabela 4. Condições para PCR da região flanqueadora (491 pb) do SNP rs2069426.

Etapa do ciclo	Temperatura (°C)	Tempo
Desnaturação inicial	95,0	10 min
Desnaturação	96,0	30 s
Anelamento dos primers	64,0	30 s
Extensão	72,0	30 s
Extensão final	72,0	10 min

40x

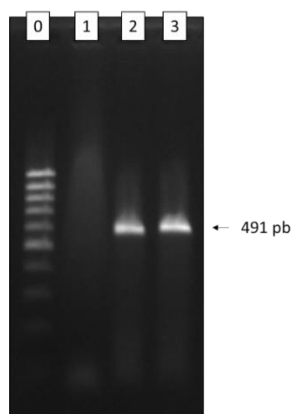


Figura 4. Gel de eletroforese da PCR da região flangeadora (491 pb) do SNP rs2069426. Legenda: 0 – ladder; 1 – branco; 2 – A1; 3 – A2.

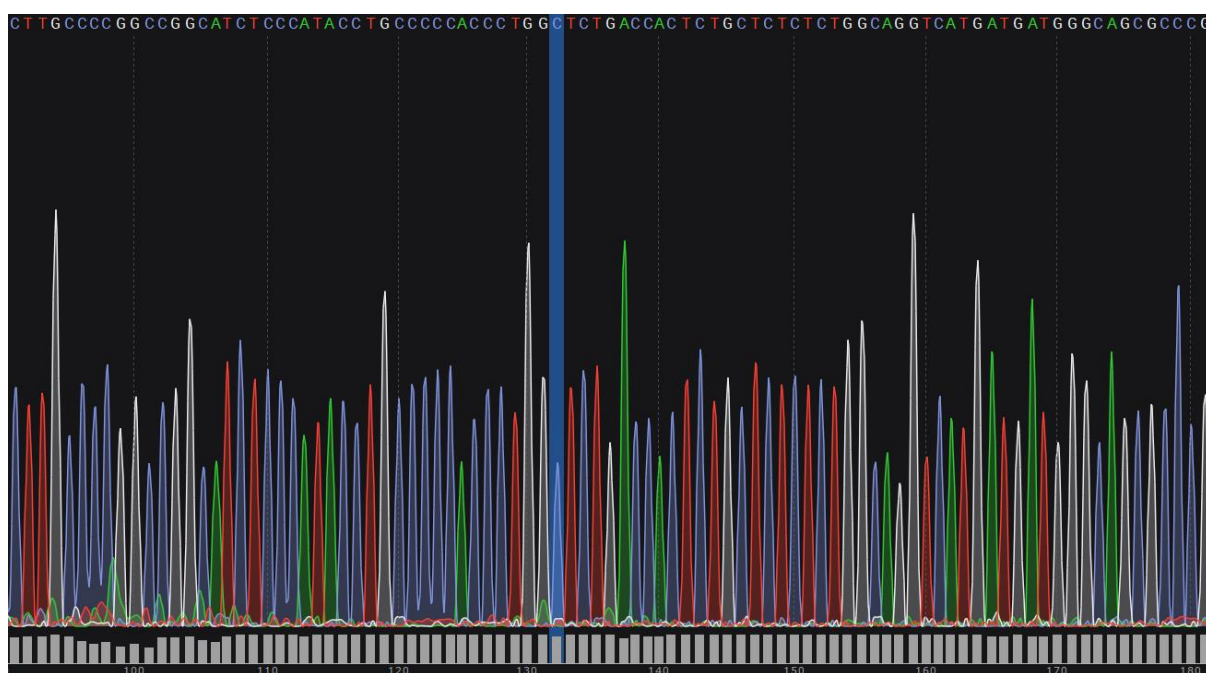


Figura 5. Cromatograma do amplicon da região flangeadora (491 pb) do SNP rs2069426. Legenda: faixa azul – alelo G do rs2069426 (fita -).

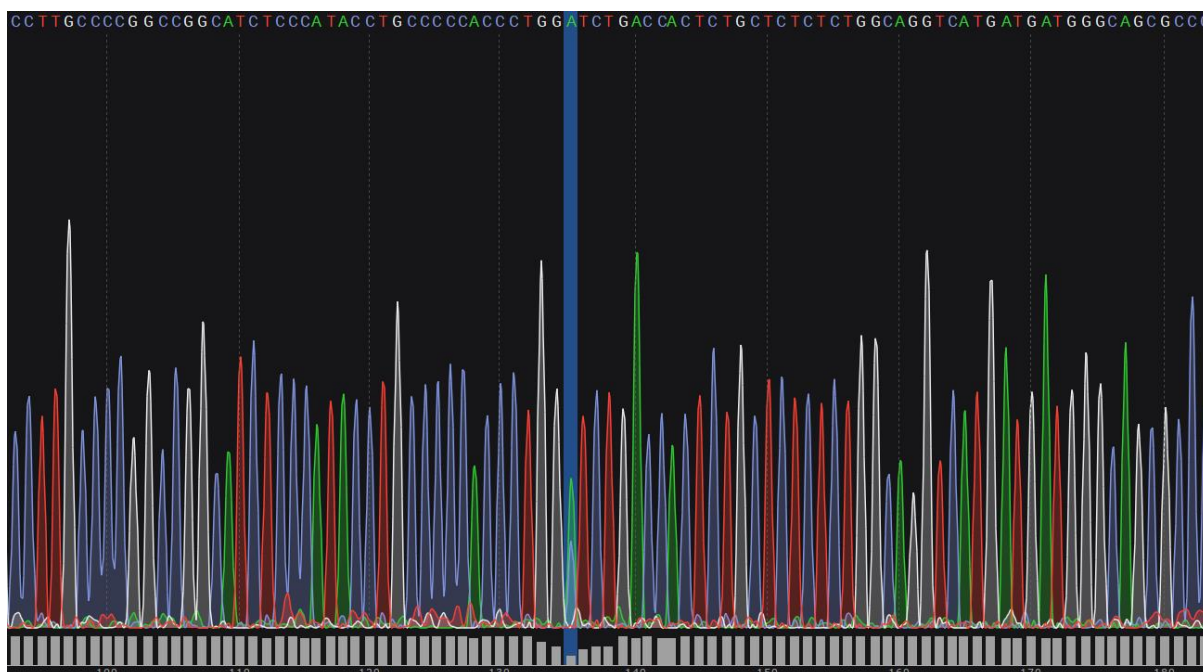


Figura 6. Cromatograma do amplicon da região flanqueadora (491 pb) do SNP rs2069426. Legenda: faixa azul – alelos T/G do rs2069426 (fita -), pico duplo, heterozigoto.



Figura 7. Alinhamento da sequência da região flanqueadora (491 pb) do SNP rs2069426 mapeada no genoma referência GRCh38. Legenda: seta vermelha – alelos T/G do rs2069426 (fita +).

Tabela 5. Genótipos das amostras LLA01-07 a partir do sequenciamento de Sanger.

Código Sequenciamento	Código Amostra	Genótipo
A1	LLA01	GG
A2	LLA02	GG
A3	LLA03	GT
A4	LLA04	GG
A5	LLA05	GG
A6	LLA06	GT
A7	LLA07	GG

4.8. SEQUENCIAMENTO REGIÃO FLANQUEADORA DE rs9290663

O gel de eletroforese do *amplicon* da região flanqueadora (226 pb) do SNP rs9290663 também mostrou bandas do tamanho esperado em relação ao que foi

desenhado *in silico* (Figura 8). As condições da reação estão mostradas na tabela 5. Assim como para o outro SNP, os cromatogramas do sequenciamento desses *amplicons* revelaram dois picos de fluorescência na posição do SNP para algumas amostras testadas, indicando genótipo heterozigoto (A/T) (Figuras 9-10). Ao todo, 3 amostras foram sequenciadas para esse teste (Figura 11) (Tabela 7).

Tabela 6. Condições para PCR da região flanqueadora (226 pb) do SNP rs9290663.

Etapa do ciclo	Temperatura (°C)	Tempo	
Desnaturação inicial	95,0	10 min	
Desnaturação inicial	96,0	30 s	} 40x
Anelamento dos primers	61,0	30 s	
Extensão	72,0	30 s	
Extensão final	72,0	10 min	

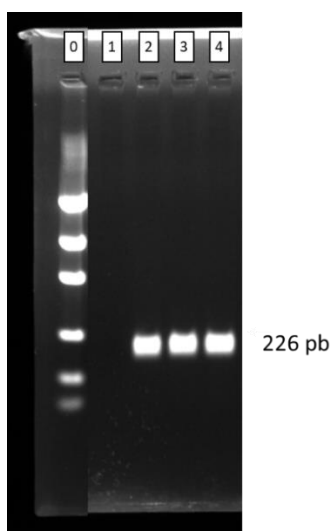


Figura 8. Gel de eletroforese da PCR da região flanqueadora (226 pb) do SNP rs9290663 (KCNMB2-AS1). Legenda: 0 – ladder; 1 – branco; 2 – A1; 3 – A2; 4 – A5.

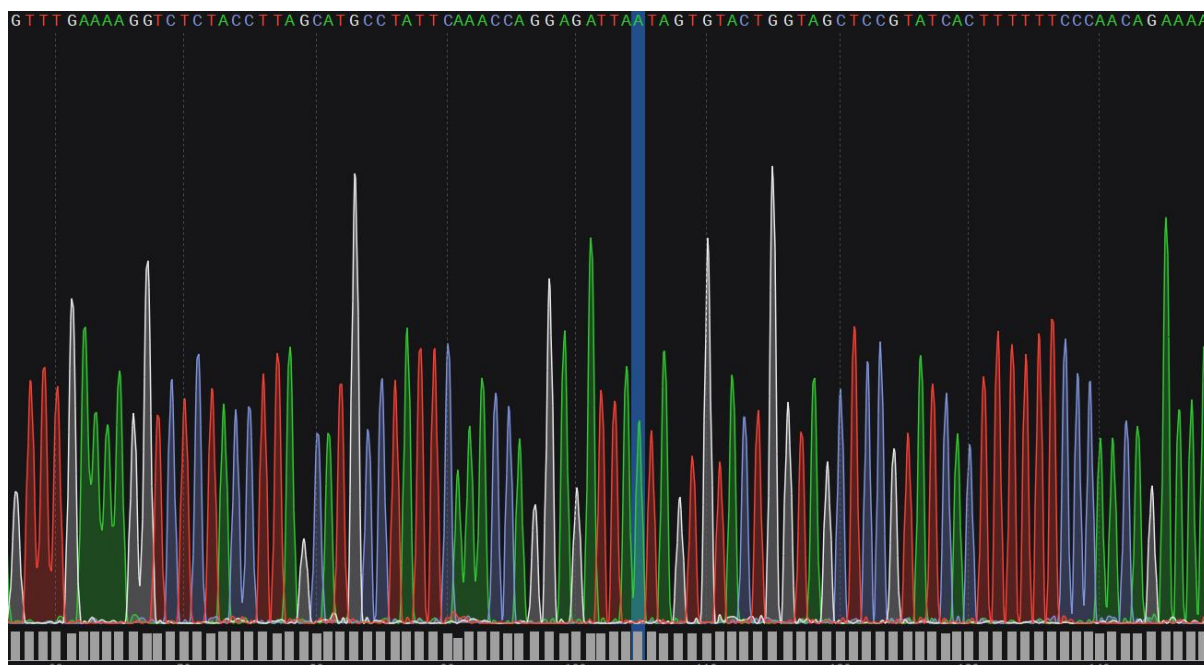


Figura 9. Cromatograma do amplicon da região flangeadora (226 pb) do SNP rs9290663. Legenda: faixa azul – alelo T do rs9290663 (fita -).

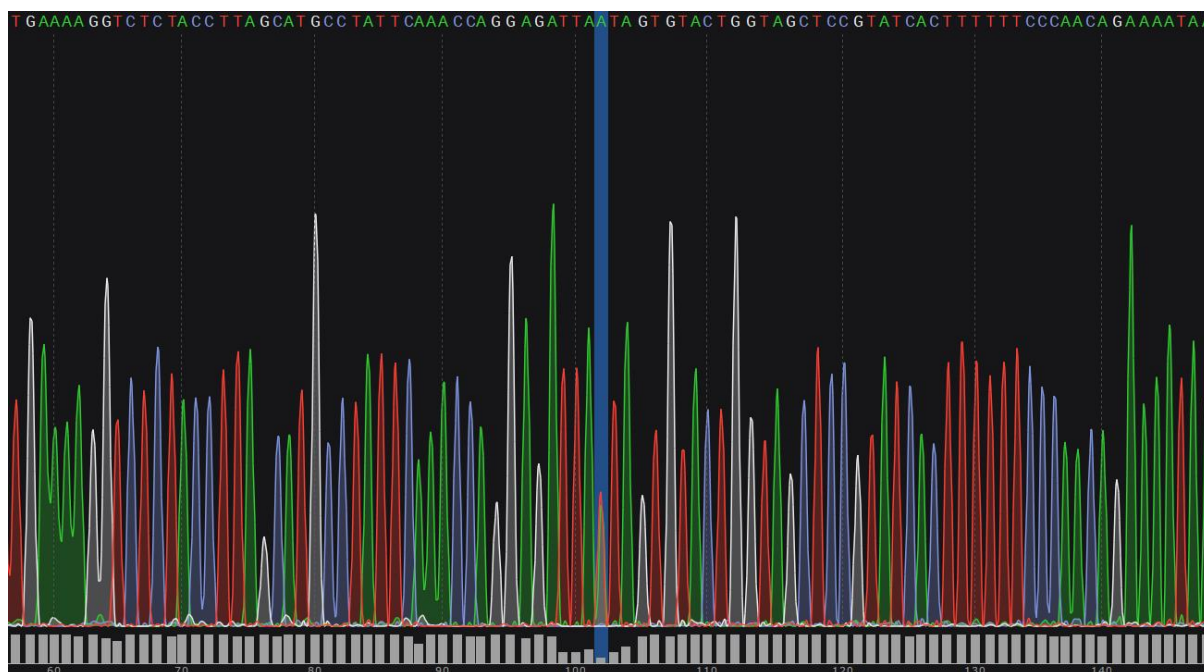


Figura 10. Cromatograma do amplicon da região flangeadora (226 pb) do SNP rs9290663. Legenda: faixa azul – alelo A/T do rs9290663 (fita -), pico duplo, heterozigoto.

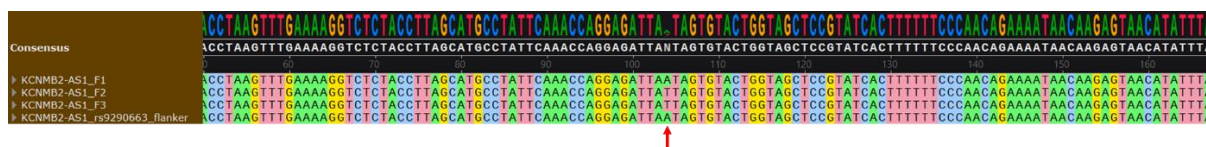


Figura 11. Alinhamento da sequência da região flanqueadora (226 pb) do SNP rs9290663 mapeada no genoma referência GRCh38. Legenda: seta vermelha – alelos A/T do rs9290663 (fita +).

Tabela 7. Genótipos das amostras LLA29-31 a partir do sequenciamento de Sanger.

Código Sequenciamento	Código Amostra	Genótipo
A1	LLA29	AA
A2	LLA30	AT
A3	LLA31	AT

4.9. PADRONIZAÇÃO DA PCR ALELO-ESPECÍFICA DE *rs2069426*

A tabela 8 apresenta as condições encontradas para a padronização da PCR alelo-específica do rs2069426. A temperatura de anelamento mais específica para a reação foi de 69°C, como mostrado no gel de agarose da figura 12.

Tabela 8. Condições para PCR alelo-específica (189 pb) do SNP rs2069426.

Etapa do ciclo	Temperatura (°C)	Tempo
Desnaturação inicial	95,0	10 min
Desnaturação inicial	96,0	30 s
Anelamento dos primers	69,0	30 s
Extensão	72,0	30 s
Extensão final	72,0	10 min

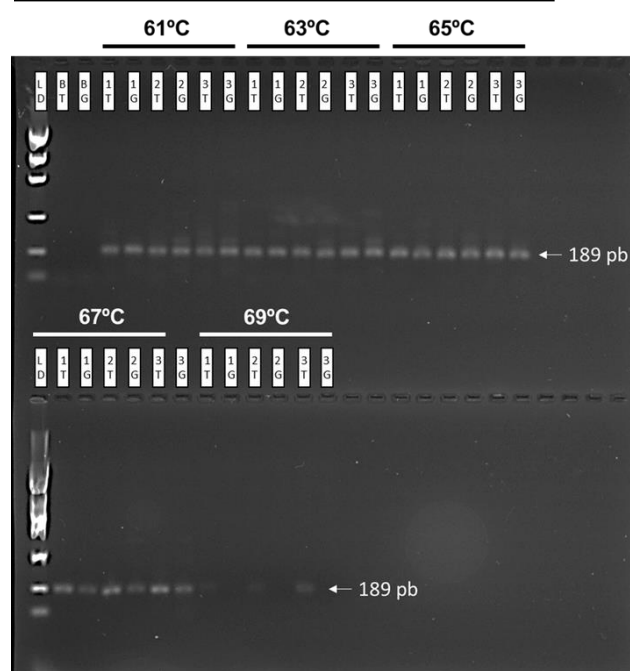


Figura 12. Gel de eletroforese da padronização da PCR alelo-específica do SNP rs2069426 (G/T).

4.10. PADRONIZAÇÃO DA PCR ALELO-ESPECÍFICA DE *rs9290663*

A padronização da PCR alelo-específica do SNP *rs9290663* foi feita pela aluna Ana Laura Beal, em colaboração com o autor. A temperatura mais específica encontrada para a reação foi de 63°C (Tabela 9).

Tabela 9. Condições para PCR alelo-específica (154 pb) do SNP *rs9290663*.

Etapa do ciclo	Temperatura (°C)	Tempo	
Desnaturação inicial	95,0	10 min	
Desnaturação inicial	96,0	30 s	} 40x
Anelamento dos primers	63,0	30 s	
Extensão	72,0	30 s	
Extensão final	72,0	10 min	

4.11. GENOTIPAGEM DE *rs2069426*

A figura 13 mostra o gel de agarose referente à PCR alelo-específica das amostras sequenciadas (mostradas nos cromatogramas das figuras 5 e 6). Como pode ser visto, os dois genótipos GG e GT foram identificados por meio da PCR. Porém, quando a PCR foi realizada exaustivamente nas amostras de LLA pediátrica, não foram obtidos os mesmos resultados. A figura 6 mostra o gel de agarose referente à PCR alelo-específica tendo como *template* o DNA das amostras de LLA pediátrica. Nenhuma banda com o tamanho esperado apareceu nos géis.

Algumas estratégias foram adotadas para driblar esse obstáculo. Primeiramente, o mix de PCR foi feito manualmente, componente por componente, para verificar se o problema estava no Mastermix do IBMP-ICC. Depois disso, verificou-se se a questão se tratava da quantidade de amostra aplicada no gel. Para isso, após a diluição das amostras (a 20 ng/μL), aquelas que não atingiram a concentração de diluição após a extração, foram divididas em duas faixas: 10-20 ng/μL e 0-10 ng/μL. As amostras da faixa 10-20 ng/μL foram ajustadas para uma quantidade de DNA 2 vezes maior na aplicação do gel, enquanto as amostras da faixa 0-10 ng/μL foram aplicadas a uma quantidade 3 vezes maior. Apesar das tentativas, nenhuma das estratégias obteve êxito.

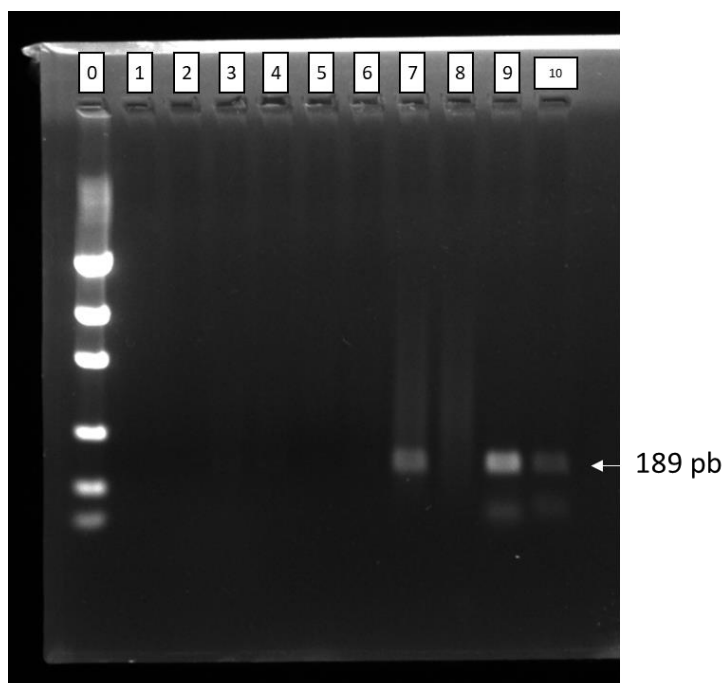


Figura 13. Gel de eletroforese da PCR alelo-específica do SNP rs2069426 (G/T) em amostras controle (amostras A1, A2, A3 e A4). Legenda: Poço 0 – marcador de peso molecular; Poço 1 – Branco G; Poço 2 – Branco T; Poço 3 – LLA03/G; Poço 4 – LLA03/T; Poço 5 – LLA04/G; Poço 6 – LLA04/T; Poço 7 – LLA05/G; Poço 8 – LLA05/T; Poço 9 – LLA06/G; Poço 10 – LLA06/T.

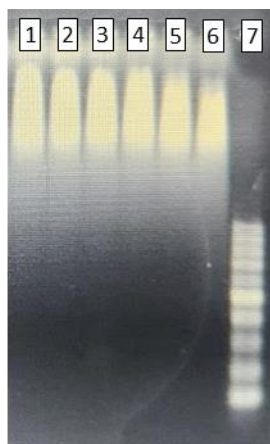


Figura 14. Gel de eletroforese da genotipagem do SNP rs2069426 (G/T) em amostras de DNA de LLA pediátrica (amostras LLA01, LLA02). Legenda: Poço 1 – Branco G; Poço 2 – Branco T; Poço 3 – LLA01/G; Poço 4 – LLA01/T; Poço 5 – LLA02/G; Poço 6 – LLA02/T; Poço 7 – marcador de peso molecular.

4.12. GENOTIPAGEM DE *rs9290663*

As figuras 15, 16 e 17 mostram os géis de agarose referentes à genotipagem do SNP rs9290663 nas amostras de DNA extraídas de sangue de paciente com LLA pediátrica. É possível perceber as bandas do tamanho esperado (154 pb), indicando os genótipos de cada indivíduo: considerando um indivíduo a cada duas bandas, uma banda à esquerda indica homozigotos AA, uma banda à direita indica homozigotos

TT e duas bandas indicam heterozigoto AT. Ao todo, 39 amostras foram genotipadas por PCR alelo-específicas para essa região (*vide* Apêndice).

O mesmo problema que ocorreu com a PCR alelo-específica do SNP do *ANRIL*, aconteceu com a PCR do SNP do *KCNMB2-AS1*, apesar de menos frequente. As mesmas estratégias foram adotadas, e da mesma forma, sem êxito.

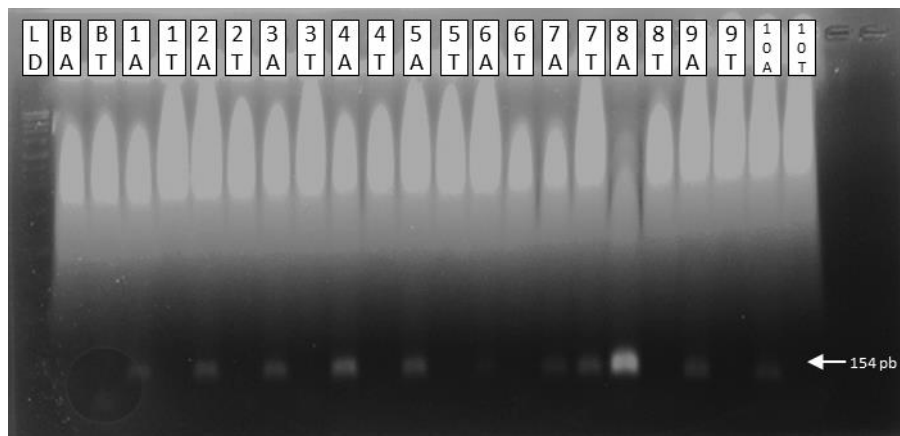


Figura 15. Gel de eletroforese da genotipagem do SNP rs9092663 (A/T) em amostras de DNA de LLA pediátrica (da esquerda para direita: ladder, branco A, branco T, amostras LLA01a, LLA02, LLA04, LLA05, LLA06, LLA07, LLA08, LLA09, LLA10).

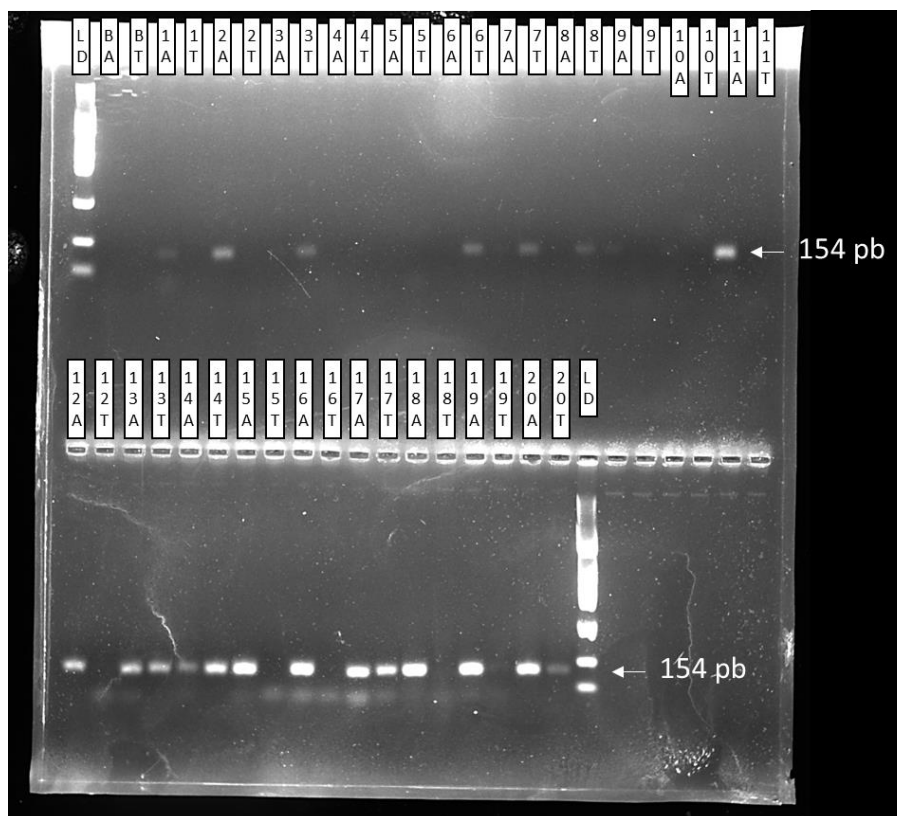


Figura 16. Gel de eletroforese da genotipagem do SNP rs9092663 (A/T) em amostras de DNA de LLA pediátrica (da esquerda para direita: ladder, branco A, branco T, amostras LLA11, LLA12, LLA13, LLA20, LLA21, LLA22, LLA26, LLA27, LLA28, LLA29, LLA30, LLA31, LLA32, LLA33, LLA34, LLA37, LLA39, LLA40).

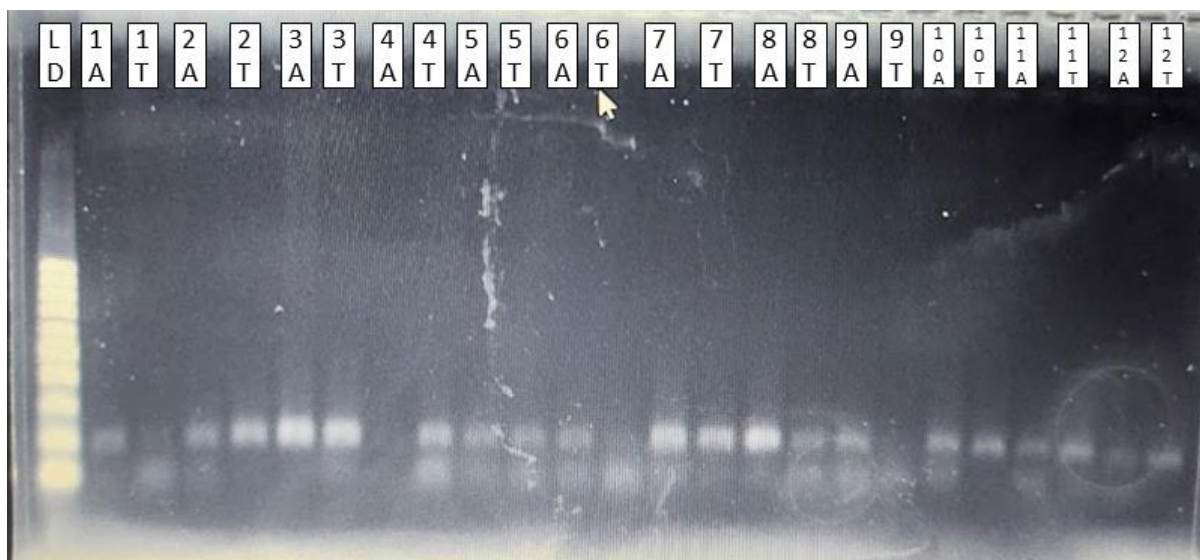


Figura 17. Gel de eletroforese da genotipagem do SNP rs9092663 (A/T) em amostras de DNA de LLA pediátrica (da esquerda para direita: ladder, amostras LLA41, LLA43, LLA44, LLA46, LLA48, LLA50, LLA51, LLA52, LLA53, LLA54, LLA55, LLA57).

5. DISCUSSÃO

Os resultados das análises computacionais mostraram regiões genômicas interessantes para validação funcional da associação estatística registrada na literatura (*ANRIL*, *KCNMB2-AS1*, *CCDC26*). Este trabalho evidencia regiões de lncRNAs com função biológica ainda não conhecida como potenciais marcadores de susceptibilidade à LLA pediátrica (*AC124014.1*, *AC016526.1*, *AC025508.1*, *AC079921.1*, *AL035446.1*, *AL355612.1*, *MMADHC-DT*) a partir da análise dos dados de 58 GWAS.

Portanto, a partir de dados públicos, o presente trabalho identificou de forma criteriosa 12 SNPs de lncRNAs associados à LLA pediátrica e selecionou os SNPs rs2069426 (no lncRNA *ANRIL*) e rs9290663 (no lncRNA *KCNMB2-AS1*) para análise detalhada. Apesar dos inúmeros desafios enfrentados durante a pesquisa, as etapas de controle de qualidade não foram negligenciadas. A quantificação das amostras de DNA extraído do sangue periférico forneceu uma base sólida para as técnicas subsequentes, a partir da exclusão de 13 amostras devido à baixa qualidade. A média de pureza das amostras (1,78) indicou uma qualidade aceitável, enquanto a diluição para 20 ng/μL e o ajuste de quantidade de DNA na aplicação do gel visou uniformidade nos experimentos.

Todo esse esforço foi necessário na tentativa de avaliar a associação desses polimorfismos em uma população brasileira, visto que tal representatividade não é predominante nos estudos atuais (MILLS & RAHAL, 2020). As duas regiões escolhidas para análise detalhada são fortes candidatas para estudos funcionais, visto que, além dos critérios usados no *workflow* metodológico, já se tem indícios de validação biológica de alguns mecanismos moleculares em que essas regiões estão envolvidas no câncer, especificamente em leucemia.

Respectivamente para os lncRNAs *ANRIL* e *KCNMB2-AS1*, os trabalhos de Chen *et al.* (2020) e Cao *et al.* (2016) conectam a lacuna entre uma associação estatística encontrada em GWAS e a função biológica da região associada. Estudos como esses são de extrema importância para a validação da associação e para o levantamento de informações relevantes para a pesquisa translacional. Afinal, o conhecimento revelado nesses trabalhos contribui para uma prática clínica mais assertiva, personalizada e menos invasiva, seja a nível de prognóstico, diagnóstico ou tratamento. Não só isso, mas esses estudos corroboram para a elucidação de mecanismos moleculares antes não explicados, como o papel regulatório de lncRNAs e como variantes nessas moléculas afetam um fenótipo tão complexo como o câncer.

As tentativas de genotipagem das duas regiões estudadas no trabalho continuarão. Assim, almeja-se a validação da associação estatística em uma coorte brasileira, para que futuramente pesquisas funcionais sejam realizadas a fim de se entender o papel dessas variantes na susceptibilidade à LLA pediátrica.

6. CONSIDERAÇÕES FINAIS

A lista de SNPs de lncRNAs associados à LLA pediátrica candidatos para validação funcional compilada aqui nesse trabalho contribui para o avanço das pesquisas oncogenéticas com os RNAs longos não codificantes. Espera-se que mais trabalhos como esse surjam, para que seja incentivada a tentativa de validação funcional das inúmeras associações estatísticas que o volume exponencial de GWAS têm depositado em bancos públicos hoje em dia.

Almeja-se, também, que esse estudo seja continuado no laboratório, pois foi mostrado aqui que essas regiões têm potencial para revelar informações importantes sobre a diversidade genética brasileira, e, futuramente, os mecanismos de ação tanto do *ANRIL* quanto do *KCNMB2-AS1* em LLA pediátrica.

Apesar de muitos desafios e obstáculos enfrentados no trabalho, é necessário destacar a importância desse estudo para minha formação acadêmica. Fazer ciência no Brasil não é fácil, e ter a oportunidade de contribuir para um trabalho desse porte foi definitivamente uma experiência valiosa. Desde as conversas com os pais dos pacientes no Hospital Oncopediátrico Erastinho, até as várias reações de PCR sem êxito. No fim, meus sinceros agradecimentos a todas as pessoas envolvidas no processo.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ALAGGIO, R., *et al.* The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms. **Leukemia**, v. 36.7, p. 1720-1748, 2022.

AMATI, B.; ALEVIZOPOULOS, K.; VLACH, J. Myc and the cell cycle. **Frontiers in Bioscience**, v. 3, n. 3, p. d250-68, 1998.

ARBER, D. A., *et al.* International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. **Blood: The Journal of the American Society of Hematology**, v. 140.11, p. 1200-1228, 2022.

ARBER, D. A.; ORAZI, A.; HASSERJIAN, R.; THIELE, J.; BOROWITZ, M. J.; LE BEAU, M. M.; BLOOMFIELD, C. D.; CAZZOLA, M.; VARDIMAN, J. W. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. **Blood**, [S. l.], v. 127, n. 20, p. 2391–2405, 2016. DOI: 10.1182/blood-2016-03-643544.

ARCHER, N. P. *et al.* Family-based exome-wide association study of childhood acute lymphoblastic leukemia among Hispanics confirms role of *ARID5B* in susceptibility. **PLoS One**, v. 12, n. 8, p. e0180488, 2017.

BRIDGES, M. C.; DAULAGALA, A. C.; KOURTIDIS, A. LNCcation: lncRNA localization and function. **Journal of Cell Biology**, v. 220, n. 2, 2021.

BUNIELLO, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. **Nucleic Acids Research**, v. 47, n. D1, p. D1005-D1012, 2019.

CAO, S. *et al.* Replication analysis confirms the association of several variants with acute myeloid leukemia in Chinese population. **Journal of cancer research and clinical oncology**, v. 142, n. 1, p. 149-155, 2016.

CHEN, L. *et al.* LncRNA CDKN2B-AS1 contributes to tumorigenesis and chemoresistance in pediatric T-cell acute lymphoblastic leukemia through miR-335-3p/TRAF5 axis. **Anti-Cancer Drugs**, 2023.

CUNNINGHAM, F. *et al.* Ensembl 2022. **Nucleic acids research**, v. 50, n. D1, p. D988-D995, 2022.

ELLINGHAUS, E. *et al.* Identification of germline susceptibility loci in *ETV6-RUNX1*-rearranged childhood acute lymphoblastic leukemia. **Leukemia**, v. 26, n. 5, p. 902-909, 2012.

FAIRLEY, Susan *et al.* The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. **Nucleic acids research**, v. 48, n. D1, p. D941-D947, 2020.

HANAHAN, D.; WEINBERG, R. A. Hallmarks of cancer: The next generation. **Cell**, [S. l.], v. 144, n. 5, p. 646–674, 2011. DOI: 10.1016/j.cell.2011.02.013. Disponível em: <http://dx.doi.org/10.1016/j.cell.2011.02.013>.

HANAHAN, D.; WEINBERG, R. A. The Hallmarks of Cancer. **Cell**, [S. l.], v. 100, n. 1, p. 57–70, 2000.

HANAHAN, Douglas. Hallmarks of Cancer: New Dimensions. **Cancer Discovery**, v. 12, n. 1, p. 31-46, 2022.

HUBER, S. *et al.* AML classification in the year 2023: How to avoid a Babylonian confusion of languages. **Leukemia**, p. 1-8, 2023.

IKEGAWA, S. A short history of the genome-wide association study: where we were and where we are going. **Genomics & informatics**, v. 10.4, p. 220, 2012.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Estimativa 2020: incidência de câncer no Brasil**. Rio de Janeiro: INCA, 2019. p. 1-120.

IRANPOUR, M.; SOUDYAB, M.; GERANPAYEH, L.; MIRFAKHRAIE, R.; AZARGASHB, E.; MOVAFAGH, A.; GHAFOURI-FARD, S. Expression analysis of four long noncoding RNAs in breast cancer. **Tumor Biology**, v. 37, n. 3, p. 2933-2940, 2016.

KENT, W. J.; SUGNET, C. W.; FUREY, T. S.; ROSKIN, K. M.; PRINGLE, T. H.; ZAHLER, A. M.; HAUSSLER, D. The human genome browser at UCSC. **Genome Research**, v. 12, n. 6, p. 996-1006, 2002.

KHOURY, J. D., *et al.* The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. **Leukemia**, p. 36.7, p. 1703-1719, 2022.

LAURENT, G. S.; WAHLESTEDT, C.; KAPRANOV, P. The Landscape of long noncoding RNA classification. **Trends in genetics**, v. 31, n. 5, p. 239-251, 2015.

MARRERO, R. J.; LAMBA, J. K. Current Landscape of Genome-Wide Association Studies in Acute Myeloid Leukemia: A Review. **Cancers**, vol. 15, n. 14, p. 3583, 2023.

MATTICK, J. S., *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. **Nature Reviews Molecular Cell Biology**, v. 24.6, p. 430-447, 2023.

MILLS, M. C.; RAHAL, C. A scientometric review of genome-wide association studies. **Communications Biology**, v. 2, n. 1, p. 1-11, 2019.

MILLS, Melinda C.; RAHAL, Charles. The GWAS Diversity Monitor tracks diversity by disease in real time. **Nature Genetics**, v. 52, n. 3, p. 242-243, 2020.

PASMANT, E., *et al.* ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. **The FASEB Journal**, v. 25.2, p. 444-448, 2011.

POPEJOY, Alice B.; FULLERTON, Stephanie M. Genomics is failing on diversity. **Nature**, v. 538, n. 7624, p. 161-164, 2016.

PRIVES, C. Signaling to p53: breaking the MDM2–p53 circuit. **Cell**, v. 95, n. 1, p. 5-8, 1998.

SHERRY, S. T.; WARD, M.; SIROTKIN, K. dbSNP - database for single nucleotide polymorphisms and other classes of minor genetic variation. **Genome Research**, v. 9, n. 8, p. 677-679, 1999.

SIEGEL, R. L.; MILLER, K. D.; JEMAL, A. Cancer statistics, 2019. **CA: a cancer journal for clinicians**, v. 69, n. 1, p. 7-34, 2019.

SONG, C.; QI, Y.; ZHANG, J.; GUO, C.; YUAN, C. CDKN2B-AS1: an indispensable long non-coding RNA in multiple diseases. **Current Pharmaceutical Design**, v. 26, n. 41, p. 5335-5346, 2020.

STATELLO, L.; GUO, C. J.; CHEN, L. L.; HUARTE, M. Gene regulation by long non-coding RNAs and its biological functions. **Nature Reviews Molecular Cell Biology**, [S. l.], v. 22, n. 2, p. 96–118, 2021. DOI: 10.1038/s41580-020-00315-9. Disponível em: <http://dx.doi.org/10.1038/s41580-020-00315-9>.

SUNG, H.; FERLAY, J.; SIEGEL, R.; LAVERSANNE, M.; SOERJOMATARAM, I.; JEMAL, A.; BRAY, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: a cancer journal for clinicians**, v. 71, n. 3, p. 209-249, 2021.

TANG, H. *et al.* Identification and function analysis of a five-long noncoding RNA prognostic signature for endometrial cancer patients. **DNA and cell biology**, v. 38, n. 12, p. 1480-1498, 2019.

TREVIÑO, L. R. *et al.* Germline genomic variants associated with childhood acute lymphoblastic leukemia. **Nature genetics**, v. 41, n. 9, p. 1001-1005, 2009.

VIJAYAKRISHNAN, J. *et al.* Identification of four novel associations for B-cell acute lymphoblastic leukaemia risk. **Nature communications**, v. 10, n. 1, p. 1-9, 2019.

VILLEGAS, V. E.; ZAPHIROPOULOS, P. G. Neighboring gene regulation by antisense long non-coding RNAs. **International Journal of Molecular Sciences**, v. 16, n. 2, p. 3251-3266, 2015.

WANG, Cheng-Hong, *et al.* LncRNA ANRIL promotes cell proliferation, migration and invasion during acute myeloid leukemia pathogenesis via negatively regulating miR-34a. **The international journal of biochemistry and cell biology**, v. 119, p. 105666, 2020.

WEINBERG, R. A. The Nature of Cancer. *In*: WEIBERG, R. A. **The Biology of Cancer**. Nova Iorque: Garland Science, 2014. p. 31-69.

WIEMELS, J. L.; WALSH, K. M.; DE SMITH, A. J.; METAYER, C.; GONSETH, S.; HANSEN, H. M.; FRANCIS, S. S.; OJHA, J.; SMIRNOV, I.; BARCELLOS, L.; XIAO, X.; MORIMOTO, L.; MCKEAN-COWDIN, R.; WANG, R.; YU, H.; HOH, J.; DEWAN, A. T.; MA, X. GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosomes 17q12 and 8q24.21. **Nature Communications**, v. 9, n. 1, p. 1-8, 2018.

WIJMENGA, C.; ZHERNAKOVA, A. The importance of cohort studies in the post-GWAS era. **Nature Genetics**, v. 50, n. 3, p. 322-328, 2018.

WILD, C. P.; WEIDERPASS, E.; STEWART, B. W. (rev.) The Global Cancer Burden. *In*: WILD, C. P.; WEIDERPASS, E.; STEWART, B. W. (org.) **World Cancer Report: Cancer Research for Cancer Prevention**. Lyon, França: International Agency for Research on Cancer, 2020. p. 15-48. Disponível em: <http://publications.iarc.fr/586>. Licença: CC BY-NC-ND 3.0 IGO.

YANG, G.; LU, X.; YUAN, L. LncRNA: a link between RNA and cancer. **Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms**, v. 1839, n. 11, p. 1097-1109, 2014.

YANG, H.; WANG, Z. Long Noncoding RNA KCNMB2-AS1 Increases ROCK1 Expression by Sponging microRNA-374a-3p to Facilitate the Progression of Non-Small-Cell Lung Cancer. **Cancer management and research**, v. 12, p. 12679, 2020.

YANG, J. J. *et al.* Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. **Nature genetics**, v. 43, n. 3, p. 237-241, 2011.

YAP, K. L.; LI, S.; MUÑOZ-CABELLO, A. M.; RAGUZ, S.; ZENG, L.; MUTJABA, S.; GIL, J.; WALSH, M. J.; ZHOU, M. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. **Molecular Cell**, v. 38, n. 5, p. 662-674, 2010.

YE, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. **BMC Bioinformatics**, v. 13, n. 1, p. 1-11, 2012.

ZHANG, F.; LUPSKI, J. R. Non-coding genetic variants in human disease. **Human molecular genetics**, v. 24.R1, p. R102-R110, 2015.

ZHANG, Y. *et al.* Long Noncoding RNA KCNMB2-AS1 Stabilized by N6-Methyladenosine Modification Promotes Cervical Cancer Growth Through Acting as a Competing Endogenous RNA. **Cell transplantation**, v. 29, p. 0963689720964382, 2020.

ZHU, J. *et al.* KCNMB2-AS1 Promotes Bladder Cancer Progression Through Sponging miR-374a-3p to Upregulate S100A10. **Frontiers in Genetics**, p. 1023, 2021.

8. APÊNDICE I

Quadro 1. Quantificação e genotipagem das amostras de DNA extraído do sangue periférico de pacientes com LLA pediátrica (amostras LLA01-48).

Amostra	[ng/μL]	260/280	Genótipo rs9290663	Genótipo rs2069426
LLA01	72,60	1,83	AA	-
LLA02	4,20	1,28	AA	-
LLA04	35,30	1,81	AA	-
LLA05	61,40	1,85	AA	-
LLA06	2,50	1,09	AA	-
LLA07	3,10	1,11	AT	-
LLA08	14,80	1,6	AA	-
LLA09	21,80	1,88	AA	-
LLA10	68,30	1,93	AA	-
LLA11	41,40	1,87	AA	-
LLA12	9,10	1,77	AA	-
LLA13	75,20	1,93	TT	-
LLA14	48,10	1,91	-	-
LLA18	5,80	2,08	-	-
LLA20	7,60	1,06	TT	-
LLA21	5,00	1,73	TT	-
LLA22	25,40	1,84	TT	-
LLA26	15,70	1,85	AA	-
LLA27	40,80	1,87	-	-
LLA28	8,90	1,67	AA	-
LLA29	48,10	1,88	AA	-
LLA30	20,80	1,8	AT	-
LLA31	10,10	1,98	AT	-
LLA32	22,90	1,83	AA	-
LLA33	35,30	1,86	AA	-
LLA34	9,90	1,8	AT	-
LLA37	23,40	1,84	AA	-
LLA39	23,90	1,85	AA	-
LLA40	5,00	1,78	AT	-
LLA41	4,90	1,81	AA	-
LLA43	9,20	1,84	AT	-
LLA44	24,80	1,85	AT	-
LLA46	8,00	1,85	TT	-
LLA48	66,30	1,79	AT	-

Quadro 2. Quantificação e genotipagem das amostras de DNA extraído do sangue periférico de pacientes com LLA pediátrica (amostras LLA48-101).

Amostra	[ng/ μ L]	260/280	Genótipo rs9290663	Genótipo rs2069426
LLA48	66,30	1,79	AT	-
LLA50	118,70	1,82	AA	-
LLA51	58,50	1,8	AT	-
LLA52	101,70	1,81	AT	-
LLA53	44,60	1,79	AA	-
LLA54	9,50	1,81	AT	-
LLA55	280,40	1,84	AT	-
LLA57	81,20	1,79	AT	-
LLA63	24,50	1,81	-	-
LLA65	23,10	1,87	-	-
LLA66	3,40	1,3	-	-
LLA67	9,00	1,85	-	-
LLA68	12,80	1,85	-	-
LLA70	59,60	1,89	-	-
LLA72	34,70	1,82	-	-
LLA74	143,90	1,99	-	-
LLA78	21,30	1,82	-	-
LLA79	13,10	1,8	-	-
LLA80	11,00	1,81	-	-
LLA81	7,80	1,85	-	-
LLA82	6,90	1,79	-	-
LLA83	3,70	1,8	-	-
LLA86	92,00	1,89	-	-
LLA87	78,50	1,9	-	-
LLA89	53,20	1,87	-	-
LLA90	67,80	1,48	-	-
LLA92	35,60	1,89	-	-
LLA95	38,40	1,8	-	-
LLA96	29,80	1,85	-	-
LLA97	15,80	1,83	-	-
LLA98	19,80	1,81	-	-
LLA99	9,00	1,86	-	-
LLA100	22,70	1,88	-	-
LLA101	18,90	1,82	-	-