

BRUNNA LETÍCIA OLIVEIRA SANTANA

ANÁLISE INTEGRATIVA DE ALTERAÇÕES MOLECULARES NOS GENES DA  
FAMÍLIA *SMYD* DE METILTRANSFERASE NOS TUMORES SÓLIDOS E  
LEUCEMIA.

Brasília

2024

UNIVERSIDADE DE BRASÍLIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MÉDICAS  
FACULDADE DE CIÊNCIAS DA SAÚDE

BRUNNA LETÍCIA OLIVEIRA SANTANA

ANÁLISE INTEGRATIVA DE ALTERAÇÕES MOLECULARES NOS GENES DA  
FAMÍLIA *SMYD* DE METILTRANSFERASE NOS TUMORES SÓLIDOS E  
LEUCEMIA.

Dissertação apresentada como requisito para a  
obtenção do título de Mestre em Ciências da Saúde  
pelo Programa de Pós-Graduação em Ciências da  
Saúde da Universidade de Brasília.

Orientador: Prof. Dr. Fábio Pittella Silva

Brasília  
2024

## **BRUNNA LETÍCIA OLIVEIRA SANTANA**

Análise integrativa de alterações moleculares nos genes da família *SMYD* de metiltransferase nos tumores sólidos e leucemia.

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre em Ciências Médicas pelo Programa de Pós-Graduação em Ciências Médicas da Universidade de Brasília

Aprovado em: 27 de Maio de 2024

### **BANCA EXAMINADORA**

Prof. Dr. Fábio Pittella Silva - (Presidente)

Universidade de Brasília

Profa. Dra. Carla Nunes de Araújo

Universidade de Brasília

Profa. Dra. Rosângela Vieira de Andrade

Universidade Católica de Brasília

Prof. Dr. Felipe Saldanha de Araújo

Universidade de Brasília

## AGRADECIMENTOS

A jornada durante o período do mestrado foi repleta de desafios tanto profissionais quanto pessoais, e o apoio de pessoas incríveis foi essencial para a conclusão dessa etapa.

Agradeço primeiramente aos meus pais, Alessandra e Carlos, por todo apoio, amor, confiança e dedicação que sempre tiveram e continuam tendo por mim. Sem vocês, nada disso seria possível!

Aos meus avós, Umbelina e Ronilson, tios e tias, por acreditarem em mim, e por todo carinho que me foi dedicado.

Ao meu namorado Luís, pelo apoio, amor, parceria e compreensão que me foi dado do início ao fim dessa trajetória. Aos meus sogros, cunhados e cunhadas, pelos momentos de alegria e leveza.

Ao professor Fábio, pela oportunidade, paciência e confiança. Sou grata pela sua compreensão e conselhos.

Ao Alan, pela paciência, amizade, ensinamentos e treinamento que foram fundamentais para o desenvolvimento desse trabalho.

À Ana e Mariana, pelos ensinamentos, incentivos, parceria, conversas e cafés! Agradeço também por compartilharem a luta em busca do western blot "publication quality", e por deixarem os dias no laboratório mais leves e agradáveis.

À equipe cancerlab, pelos ensinamentos e momentos de descontração.

E, por fim, agradeço aos professores Carla Araújo, Rosângela de Andrade e Felipe Saldanha por comporem minha banca e pela contribuição a esse trabalho, aos órgãos de fomento e ao programa de Pós-graduação de Ciências Médicas da Universidade de Brasília.

# CAPÍTULO 1

## RESUMO

A família *SMYD*, que contém os domínios conservados SET e MYND, é composta por cinco membros que possuem função de lisina metiltransferase. Um número crescente de evidências indica o envolvimento dos genes da família *SMYD* em diferentes tipos de câncer. Aqui, o objetivo foi investigar as alterações genéticas mais comuns da família *SMYD* entre os tumores sólidos mais frequentes e determinar seu potencial uso como biomarcadores. Foi realizado uma análise integrativa de várias plataformas sobre as mutações mais comuns, alterações no número de cópias (CNA) e expressão de mRNA. Dados de coortes de carcinoma de mama invasivo (BRCA), adenocarcinoma de próstata (PRAD), carcinoma colorretal (CRC), adenocarcinoma de estômago (STAD), adenocarcinoma de pulmão (LUAD), carcinoma de células escamosas de pulmão (LUSC), carcinoma endometrial de corpo uterino (UCEC) e carcinosarcoma uterino (UCS) disponíveis no Atlas do Genoma do Câncer (TCGA) e no cBioPortal foram analisados.

Foi descoberto que os genes *SMYD* têm uma frequência geral de mutação mais baixa entre os tumores examinados. A maioria são mutações missense espalhadas entre seus domínios. Diferentes mutações foram identificadas no *SMYD1* (78), *SMYD2* (33), *SMYD3* (44), *SMYD4* (39) e *SMYD5* (44). É interessante notar que as mutações em todos os genes *SMYD* foram encontradas com mais frequência em UCECs. Portanto, nenhuma mutação foi detectada com frequência na mesma coorte e nenhuma mutação específica teve poder de estratificação suficiente para ser usada como biomarcador nos cânceres examinados.

Em relação ao CNA, a alteração de ganho de baixo nível foi a mais comum entre os genes, com exceção do *SMYD4*, que teve a perda heterozigótica como a alteração mais frequente em todos os tumores analisados.

Na expressão do mRNA, cada membro da família *SMYD* teve um padrão de expressão distinto. O *SMYD1* se dividiu entre baixa expressão e nível basal, enquanto o *SMYD2* e *SMYD3* tiveram alterações dependendo do tumor analisado, mas a maioria tendeu ao aumento da expressão; o *SMYD4* mostrou uma grande tendência à baixa expressão, sugerindo uma ação de supressor tumoral; o *SMYD5* se apresentou com alta e baixa expressão, a depender do tumor avaliado. Com relação à análise de sobrevivência, *SMYD3* e *SMYD4* em STAD, *SMYD3* em BRCA, *SMYD2* e *SMYD4* em LUAD apresentaram diferenças no tempo de vida, dependendo do nível de expressão. Portanto, os resultados mostram que os membros da família *SMYD* têm funções diferentes dependendo do tumor, e podem agir como oncogene e supressor tumoral, o que torna ainda mais evidente a importância de uma melhor compreensão do envolvimento desses genes durante a carcinogênese e sua utilização como potenciais biomarcadores moleculares.

**Palavras-chave:** Família *SMYD*; Carcinogênese, Tumores sólidos; Epigenética; Lisina metiltransferase.

## ABSTRACT

The SMYD family, which contains the conserved SET and MYND domains, comprises five members that have a lysine methyltransferase function. A growing number of evidences indicate the involvement of *SMYD* family genes in different types of cancer. Here, the aim was to investigate the most common genetic alterations of the *SMYD* family among the most frequent solid tumors and determine their potential use as biomarkers. A multi-platform integrative analysis was performed on the most common mutations, copy number alterations (CNA) and mRNA expression. Data from cohorts of invasive breast carcinoma (BRCA), prostate adenocarcinoma (PRAD), colorectal carcinoma (CRC), stomach adenocarcinoma (STAD), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), uterine corpus endometrial carcinoma (UCEC) and uterine carcinosarcoma (UCS) available on The Cancer Genome Atlas (TCGA) and cBioPortal were analyzed.

*SMYD* genes were found to have a lower overall mutation frequency among the tumors examined. Most are missense mutations scattered between their domains. Different mutations were identified in *SMYD1* (78), *SMYD2* (33), *SMYD3* (44), *SMYD4* (39) and *SMYD5* (44). It is interesting to note that mutations in all *SMYD* genes were found more frequently in UCECs. Therefore, no mutation was detected frequently in the same cohort and no specific mutation had sufficient stratification power to be used as a biomarker in the cancers examined.

With regard to CNA, the low-level gain alteration was the most common among the genes, with the exception of *SMYD4*, which had heterozygous loss as the most frequent alteration in all the tumors analyzed.

With regard to mRNA expression, each member of the *SMYD* family had a distinct expression pattern. *SMYD1* was divided between downregulated expression and basal level, while *SMYD2* and *SMYD3* had alterations depending on the tumor analyzed, but most tended to upregulation; *SMYD4* showed a great tendency to downregulation, suggesting tumor suppressor action; *SMYD5* was both upregulated and downregulated, depending on the tumor evaluated. With regard to survival analysis, *SMYD3* and *SMYD4* in STAD, *SMYD3* in BRCA and *SMYD2* and *SMYD4* in LUAD showed differences in survival time, depending on the level of expression. Therefore, the results show that *SMYD* family members have different functions depending on the tumor, and can act as oncogene and tumor suppressor, which makes it even more evident the importance of a better understanding of the involvement of these genes during carcinogenesis and their use as potential molecular biomarkers.

**Keywords:** SMYD family; Carcinogenesis, Solid tumors; Epigenetics; Lysine methyltransferase

## LISTA DE FIGURAS

### Capítulo 1

Figura 1 – *Hallmarks* do câncer

Figura 2 – Representação da histologia mamária

Figura 3 – Representação de células do ducto prostático

Figura 4 – Representação da região colorretal e seus fatores de risco

Figura 5 – Representação esquemática pulmonar

Figura 6 – Representação esquemática gástrica

Figura 7 – Representação esquemática do útero

Figura 8 – Representação da mono, di e trimetilação em resíduos de lisina

Figura 9 – Representação dos principais domínios da família SMYD

Figura 10 – Delineamento experimental

Figura 11 – Total de mutações somáticas em domínios conservados da família SMYD nos tumores sólidos

Figura 12 – Análise de sobrevida global da família SMYD referente a mutação

Figura 13 – Frequência de alterações do número de cópias na família SMYD nos tumores sólidos mais incidentes

Figura 14 – Análise de sobrevida global da família SMYD referente ao número de cópias alteradas

Figura 15 – Análise de sobrevida livre de doença da família SMYD referente ao número de cópias alteradas

Figura 16 – Expressão do mRNA da família SMYD em tumores sólidos

Figura 17 – Análise de sobrevida da família SMYD referente ao mRNA

Figura 18 – Correlação da expressão do mRNA da família SMYD entre tumores sólidos

## Capítulo 2

Figura 1 – Representação do processo hematopoiético

Figura 2 – Delineamento experimental

Figura 3 – Configuração da placa de 6 poços para processo de transdução celular

Figura 4 – Configuração da placa de 6 poços para processo de seleção de células transduzidas

Figura 5 – Representação da diluição de matriz seriada

Figura 6 – Representação das guias do *SMYD5* para knockout gênico

Figura 7 – Frequência de alterações do número de cópias do DNA presentes na família SMYD em LLA- B

Figura 8 – Expressão do mRNA da família SMYD em medula óssea saudável e LLA-B

Figura 9 – Expressão do *SMYD5* em amostras de LLA comparado a medula óssea não neoplásica.

Figura 10 – qPCR para análise de expressão do *SMYD5* em linhagem celular de LLA-B e PBMC

Figura 11 – PCR e Western blot para confirmação de expressão da Cas9

## LISTA DE TABELA

### Capítulo 1

Tabela 1 – Número de estudos e pacientes selecionados

Tabela 2 – Total de diferentes tipos de mutações somáticas encontradas nos membros da família *SMYD* em oito tumores sólidos analisados.

Tabela 3 – Frequência de mutação em cada membro da família *SMYD* entre os oito tipos de tumores.

Tabela 3 – Correlação entre a expressão de CNA e mRNA no câncer de mama

### Capítulo 2

Tabela 1 – Sequência de primers

Tabela 2 – Sequência de sgRNA

Tabela 3 – Sequência de primers de amplificação e sequenciamento

Tabela 4 – Correlação entre CNA e expressão do mRNA do *SMYD5*

## LISTA DE ABREVIATURAS

SET	Su(var)3-9, Enhancer-of-zeste e Trithorax
MYND	Myeloid, Nervy e DEAF-1
CNA	Copy number alteration (Número de cópias alteradas)
mRNA	RNA mensageiro
BRCA	Breast invasive carcinoma (Câncer de mama invasivo)
PRAD	Prostate adenocarcinoma (Adenocarcinoma de próstata)
CRC	Colorectal cancer (Câncer colorretal)
STAD	Stomach adenocarcinoma (Adenocarcinoma de estômago)
LUAD	Lung Aadenonarcinoma (Adenocarcinoma de pulmão)
LUSC	Lung squamous cell carcinoma (Câncer de pulmão de células escamosas)
UECE	Uterine corpus endometrial carcinoma (Carcinoma endometrial do corpo uterino)
UCS	Uterine Carcinosarcoma (Carcinossarcoma uterino)
TCGA	The Cancer Genome Atlas
Rb	Retinoblastoma
VEGF	Vascular endothelial growth fator (Fator de Ccrescimento do endotélio vascular)
DNA	Ácido desoxirribonucleico
HMT	Histona metiltransferase
ER	Estrogen receptors (Receptor de estrogênio)
PR	Progesterone receptors (Receptor de progesterona)
HER2	Human epidermal growth factor receptor 2 (Receptor tipo 2 do fator de crescimento epidérmico humano)
TNBC	Triple negative breast cancer (Câncer de mama triplo Negativo)

AR	Androgen receptor (Receptor de androgênio)
PSA	Prostate-specific antigen (Antígeno prostático específico)
SCLC	Small cell lung cancer (Câncer de pulmão de células pequenas)
NSCLC	Non-small cell lung cancer (Câncer de pulmão de células não pequenas)
TTF-1	Thyroid transcription factor 1 (Fator de transcrição da tireoide 1)
KRT-7	Keratin 7 (Queratina 7)
CagA	Cytotoxin-associated gene A (Citotoxina associada ao gene A)
KMT	Lysine methyltransferase (Metiltransferase de lisina)
SAM	S-adenosil-L-metionina
PcG	Polycomb complexes (Complexo Polycomb)
TrxG	Trithorax-group proteins (Grupo de proteínas Trithorax)
HSP90	Heat shock protein 90 (Proteína de choque térmico 90)
PTEN	Phosphatase and tensin homolog (Fosfatase homóloga à Tensina)
VEGFR1	Vascular endothelial growth factor receptor 1 (Receptor 1 do fator de crescimento endotelial vascular)
HDAC	Histone deacetylase (Histona desacetilase)
PDGFR-A	Platelet-derived growth factor receptor A (Receptor A do fator de crescimento derivado de plaquetas)
MiR-1307-3p	Micro RNA 1307-3p
HP1	Heterochromatin protein 1 (Proteína da heterocromatina 1)
OS	Overall survival (Sobrevida global)
PFS	Progression free survival (Sobrevida livre de progressão da doença)
DSF	Disease free survival (Sobrevida livre de doença)

CTH	Célula tronco hematopoiética
CTL	Célula tronco leucêmica
MEC	Matriz extracelular
LMC	Leucemia mieloide crônica
LMA	Leucemia mieloide aguda
LLC	Leucemia linfóide crônica
LLA-B	Leucemia linfóide aguda de célula B
RNA	Ácido ribonucleico
cDNA	DNA complementar
qPCR	PCR quantitativa
C <sub>T</sub>	Ciclo treshold
PBMC	Célula mononuclear do sangue periférico
MOS	Medula óssea saudável

# Sumário

## CAPÍTULO 1

<b>1. INTRODUÇÃO</b> .....	16
1.1. O que é o câncer? .....	16
1.2. Biologia do câncer .....	17
1.3. Genética e epigenética do câncer .....	20
1.4. O câncer no Brasil .....	23
1.4.1. Câncer de mama .....	23
1.4.2. Câncer de próstata .....	25
1.4.3. Câncer colorretal .....	26
1.4.4. Câncer de pulmão .....	28
1.4.5. Câncer de estômago .....	29
1.4.6. Câncer de útero .....	31
1.5. Metiltransferases e o câncer .....	33
1.6. Família SMYD de metiltransferases .....	34
1.6.1. <i>SMYD1</i> .....	35
1.6.2. <i>SMYD2</i> .....	36
1.6.3. <i>SMYD3</i> .....	37
1.6.4. <i>SMYD4</i> .....	38
1.6.5. <i>SMYD5</i> .....	39
<b>2. OBJETIVOS</b> .....	40
2.1. Objetivos gerais .....	40
2.2. Objetivos específicos .....	40
<b>3. METODOLOGIA</b> .....	41
3.1. Delineamento experimental .....	41
3.2. Análise <i>in silico</i> .....	41
3.3. Mapa de calor .....	43

3.4. Análise de sobrevivência .....	44
3.5. Estatística.....	44
<b>4. RESULTADOS.....</b>	<b>44</b>
4.1. Perfil mutacional .....	44
4.2. Avaliação do número de cópias alteradas.....	48
4.3. Perfil de expressão do mRNA .....	53
4.3.1. Correlação do mRNA entre os membros da família SMYD .....	57
<b>5. DISCUSSÃO .....</b>	<b>60</b>
<b>6. CONCLUSÃO .....</b>	<b>66</b>
<b>7. REFERÊNCIAS .....</b>	<b>68</b>
<b>1. INTRODUÇÃO .....</b>	<b>79</b>
1.1. O que é leucemia?.....	79
1.2. Leucemia Linfóide Aguda de células B .....	82
1.2.1. <i>SMYDs</i> e a LLA-B.....	84
<b>2. MATERIAIS E MÉTODOS.....</b>	<b>85</b>
2.1. Delineamento experimental.....	85
2.2. Base de dados online .....	85
2.3. Cultivo celular.....	85
2.4. Extração de DNA.....	86
2.5. Extração de RNA e síntese de cDNA.....	86
2.6. PCR convencional e PCR quantitativa (qPCR).....	87
2.6.1. Análise de dados da qPCR .....	87
2.7. Produção de lentivirus e Transfecção de linhagem celular .....	88
2.7.1. Transdução lentiviral .....	88
2.7.2. Separação clonal.....	90
2.7.3. Transfecção celular utilizando sgRNA.....	91
2.8. Sequenciamento de Sanger.....	92
2.9. Western Blot.....	93

<b>3. RESULTADOS</b> .....	94
3.1. Avaliação do número de cópias alteradas e expressão do mRNA.....	94
3.1.1. Correlação entre CNA e mRNA do <i>SMYD5</i> .....	96
3.2. Nível de expressão do <i>SMYD5</i> em linhagens celulares de LLA-B .....	96
3.3. Seleção de clone celular com expressão estável de Cas-9 .....	97
<b>4. DISCUSSÃO</b> .....	98
<b>5. CONCLUSÃO</b> .....	101
<b>6. PERSPECTIVAS</b> .....	101
<b>7. BIBLIOGRAFIA</b> .....	102
<b>8. ANEXO I</b> .....	106

## 1. INTRODUÇÃO

### 1.1. O que é o câncer?

O câncer é uma doença caracterizada pelo descontrole da proliferação celular podendo se disseminar para outras partes do corpo (1). Quando a proliferação celular descontrolada leva a formação de um tumor, o mesmo pode ser classificado de duas formas a depender de sua agressividade: benigno ou maligno. O tumor benigno tem como característica o crescimento local e não ocorre a invasão em tecidos adjacentes e, quando removido, não costuma crescer novamente. Já o tumor maligno tem como característica a capacidade de invadir tecidos adjacentes, e se espalhar para outros locais do corpo para formar novos tumores, processo esse conhecido por metástase. Quando removido, pode vir a crescer novamente no mesmo local e, por isso, os tumores malignos costumam ser mais agressivos e trazer riscos a vida do paciente (2).

Ao analisar o tecido tumoral, percebe-se que há diferença em sua disposição quando comparado ao tecido normal porque as células perdem sua capacidade de organização. Com isso, através de análises histológicas, é possível classificar o tipo tumoral de acordo com o seu tecido de origem. Dessa forma, o tipo mais comum é o carcinoma que surge de células epiteliais que, a depender do modelo e função específica da célula, pode ser classificados ainda em dois subtipos principais: adenocarcinoma, que surge de células epiteliais especializadas em secretar substâncias, ou carcinoma de células escamosas, que é oriundo de células epiteliais de revestimento que serve como uma camada de proteção tanto da superfície corporal quanto de órgãos e cavidade interna (1,2).

Entre os tumores do tipo não epitelial que são oriundos do tecido conjuntivo, encontra-se o sarcoma, que engloba cerca de 1% dos casos de câncer que são formados em tecidos moles e ósseos, como por exemplo os músculos, vasos sanguíneos e linfáticos, e tecido fibroso. (1,2) Existe também o câncer originado de células hematopoiéticas que incluem as células do sistema imune, sendo as leucemias, que afetam as linhagens celulares hematopoiéticas na medula óssea, e os linfomas, que afetam as linhagens celulares linfoides T e B presentes no linfonodo.

Um outro tipo não epitelial é classificado de neuroectodermal, que surge em células que compõem o sistema nervoso central e periférico e os tumores são subdivididos e chamados de glioblastomas, gliomas, neuroblastomas, schwannomas e meduloblastomas (2).

## 1.2. Biologia do câncer

As células cancerígenas adquirem e acumulam diversas mudanças estruturais que levam a desregulação de vários fatores responsáveis pelo bom funcionamento celular. Segundo Hanahan e Weinberg, existem algumas características que são encontradas desreguladas com frequência em células cancerígenas denominados de *hallmarks* do câncer, que são características adquiridas pela célula durante o processo de carcinogênese que possibilitam a sobrevivência, proliferação e disseminação da célula cancerígena. Os *hallmarks* propostos englobam a sustentação de sinais proliferativos, evasão de supressores de crescimento, resistência a morte celular, permissão da imortalidade replicativa, indução ou acessibilidade ao sistema vascular, ativação da invasão e metástase, escape da destruição pelas células imunes e desregulação do metabolismo celular (Figura 1) (3,4).

A proliferação crônica é uma característica essencial para manter a sustentação dos sinais proliferativos nas células cancerígenas. Esses sinais que permitem a proliferação são em sua grande maioria convertidos através de fatores de crescimento que vão se ligar a receptores presentes na superfície celular que normalmente possuem domínios de tirosina quinase intracelular (3), mantendo dessa forma os sinais de transdução sempre ativos e afetando os processos de transcrição gênica. Mutações nesses receptores podem também levar a uma ativação aberrante na via de sinalização intracelular que pode acarretar em alguns distúrbios como câncer, inflamação e angiogênese (5).

As células cancerígenas também podem estimular sua proliferação descontrolada ao produzir seus próprios fatores de crescimento obtendo uma estimulação autócrina de proliferação, ou através do estímulo de células normais que se encontram associadas ao estroma tumoral, que fornecem fatores de crescimento a células cancerígenas através de uma estimulação parácrina (3,6).



**Figura 1:** *Hallmarks* do câncer. Representação das características adquiridas pela célula durante o desenvolvimento tumoral. (Adaptado de Hanahan D. *Hallmarks of Cancer: New Dimensions*. *Cancer Discov.* 2022).

Outra característica comum em células cancerígenas é o processo de evasão de genes supressores tumorais, que são responsáveis pelo controle da proliferação celular. Os genes retinoblastoma (RB) e TP53 são exemplos bem estabelecidos de supressores tumorais que frequentemente encontram-se desregulados em diversos tumores, o que leva ao descontrole e aumento da taxa de proliferação celular (3,7).

A resistência a morte celular é outra característica estratégica que a célula tumoral adquire como forma de burlar a morte celular programada e, uma das vias utilizadas para tal feito envolve a perda do TP53 que acarreta na falha de uma das vias de indução a apoptose (3), e a desregulação de fatores anti e pró-apoptóticos como, por exemplo, o aumento da expressão do Bcl-2 e Bcl-xL, que são reguladores anti apoptóticos, e a downregulação de Bax, Bim, Noxa e Puma, que são fatores pró-apoptóticos, auxiliam na sobrevivência da célula cancerígena contribuindo para formação e progressão tumoral, além de terem implicações na resistência aos tratamentos oncológicos (8,9).

Para que haja a formação tumoral, sabe-se que é necessário o descontrole da proliferação celular, e uma das características que permite essa alteração é a

capacidade permissiva da imortalidade replicativa adquirida pela célula cancerígena. Uma célula normal possui uma quantidade limitada de replicação e, quando essa característica é perdida, a célula passa a se replicar de forma ilimitada levando a formação do tumor. Nesse processo de desregulação, duas barreiras que servem para impedir a proliferação são afetadas, sendo a senescência, que é um processo pelo qual a célula para de se proliferar mantendo sua viabilidade, e a crise celular, que engloba a morte celular programada. Esse descontrole ocorre porque o telômero passa a ser produzido e adicionado ao cromossomo através da reativação da telomerase, o que mantém o cromossomo protegido através das múltiplas divisões celulares não levando ao encurtamento dos telômeros, que em condições normais, ativaria a senescência e morte celular e, por isso, a célula cancerígena adquire capacidade de continuar proliferando indefinidamente (3,7,10).

Para que o tumor consiga passar de 1 milímetro, é necessário que o mesmo consiga obter sustentação de vasos sanguíneos, portanto, a angiogênese é induzida bem cedo, durante o início do desenvolvimento neoplásico com a liberação de fatores angiogênicos, sendo o mais comum o fator de crescimento do endotélio vascular (VEGF), que pode se tornar *upregulado* em situação de hipoxia e através de sinais oncogênicos, situação comum em um microambiente tumoral. Por essa razão, o VEGF é um fator bem estabelecido e usado inclusive como alvo terapêutico contra o desenvolvimento da neovascularização tumoral. Dessa forma, a angiogênese contribui também para a progressão, invasão e metastização do tumor (3,7,11).

O processo de invasão e metástase é a causa mais comum de óbito em pessoas com câncer. As células cancerígenas adquirem essa habilidade através de alterações estruturais e na capacidade de aderência a outras células, e um dos motivos desse acontecimento se dá pela perda de uma molécula chamada E-caderina, que é fundamental para a aderência celular. O programa de transição epitélio-mesenquimal também auxilia na indução da invasão e metástase por transformar as células epiteliais, possibilitando a sua circulação através dos vasos sanguíneos e linfáticos e, ao invadir e encontrar um local propício para se estabelecer, a célula pode passar pelo processo inverso passando pela transição mesenquimal-epitelial, formando assim uma nova colônia de células cancerígenas em outro órgão com característica do tumor primário (3,12).

Outro aspecto que torna possível a sobrevivência e desenvolvimento da célula cancerígena é a capacidade que a mesma adquire de burlar o sistema imune e passar despercebida, podendo se estabelecer e iniciar a formação tumoral. As células tumorais conseguem tal feito porque as mesmas liberam fatores como citocinas imunossupressoras, prostaglandinas e fator de crescimento do endotélio vascular, que são responsáveis por reorientar a resposta imune para a formação de um ambiente propício ao crescimento, ao eliminar a ação efetiva de células T citotóxicas e *Natural Killer* (3,7,13).

A reprogramação da energia metabólica é uma característica que tem se mostrado essencial no desenvolvimento tumoral. Um dos fatores dessa reprogramação metabólica é o uso da glicose como fonte energética mesmo quando a célula cancerígena encontra-se na presença de oxigênio, e essa mudança é favorável por gerar metabólitos que funcionam como matéria-prima para o crescimento e divisão celular, por gerar nucleosídeos e aminoácidos que são constituintes que facilitam a formação de macromoléculas e organelas (3). Ao usar a glicose como fonte de energia ocorre também a produção de lactato, que é um componente que auxilia na acidificação do microambiente tumoral que pode ter como consequência a aquisição de resistência a fármacos usados para tratamento oncológico (14).

### 1.3. Genética e epigenética do câncer

As células cancerígenas se desenvolvem a partir de alterações tanto genéticas quanto epigenéticas. Em geral, as alterações genéticas são ocasionadas por mutações, instabilidade no genoma e variação no número de cópia de um gene. Já as alterações epigenéticas englobam a metilação no DNA, modificação em histonas e perda do imprinting, porém essas mudanças modificam a expressão gênica sem que haja alteração na sequência do código genético (15).

A nível genético, o câncer pode iniciar através de mutação em célula germinativa e, dessa forma, todas as células do organismo são impactadas com a mutação, e esse caso é conhecido como síndrome do câncer hereditário. Uma outra forma é o desenvolvimento do câncer de forma esporádica, sendo esse a maioria dos casos relatados. Nesse processo esporádico, a mutação ocorre apenas em uma célula somática que, por sua vez, consegue se multiplicar e formar a massa tumoral.

As mutações genéticas responsáveis pelo surgimento e progressão tumoral ocorrem em sua grande maioria em duas classes de genes: os oncogenes e os genes supressores tumorais (16).

O oncogene é a forma mutada de um proto-oncogene. O proto-oncogene é a forma normal de genes que possuem como função a codificação de proteínas responsáveis pelo crescimento e sobrevivência celular. Para a transformação de um proto-oncogene em um oncogene, basta que ocorra uma mutação em apenas um alelo do gene, o que propicia um ambiente solícito a malignidade através da estimulação da proliferação ou a inibição da apoptose (16). Essa mutação desestabiliza o proto-oncogene ao afetar a estrutura da proteína ou desregular a expressão proteica.

A ativação do proto-oncogene em oncogene se dá através da translocação cromossômica que levará um local do cromossomo não transcrito a outro que será transcrito, possibilitando a formação de proteínas anormais; da mutação de ponto, que desencadeia a mudança de um aminoácido do proto-oncogene levando a formação de um oncogene e, por fim, através da amplificação gênica, que eleva o número de cópias de um oncogene acarretando em um aumento da produção de oncoproteínas (17).

Um exemplo bem estabelecido de um proto-oncogene que é encontrado amplamente mutado em diversos tumores pertence à família MYC, que desempenha funções na transformação, desdiferenciação, proliferação e imortalização celular. Portanto, a ativação do MYC no tumor estimula muitos *hallmarks* do câncer, possibilitando o desenvolvimento tumoral (17,18).

Os genes supressores tumorais são classificados em controladores e de manutenção, e contribuem com o câncer quando há a perda dos dois alelos do gene. Os controladores são responsáveis pelo controle dos mecanismos que englobam o crescimento celular, que leva ao bloqueio do desenvolvimento tumoral. Já os de manutenção são responsáveis pela proteção da integridade genômica, e a perda de sua função ocasiona no acúmulo de mutações secundárias em oncogenes e em genes supressores tumorais controladores que, juntos, levam a progressão tumoral (16,17).

De modo geral, os genes supressores tumorais estão envolvidos em vias responsáveis pela supressão da divisão celular, indução da apoptose, reparo ao dano no DNA e inibição da metástase (19) e, por essa questão, quando silenciados contribuem com a iniciação e o desenvolvimento do tumor. Um supressor tumoral bastante conhecido por ser encontrado mutado em cerca de 50% dos tumores diagnosticados é o p53, que quando alterado é responsável por levar ao descontrole da regulação da expressão de genes que controlam a proliferação, senescência, reparo do DNA e a morte celular (20).

Em relação a epigenética, os processos mais estudados são a metilação no DNA e as modificações pós-traducionais em histonas que, quando alterados, podem levar ao desenvolvimento tumoral. A metilação no DNA tem uma importante função regulatória nos processos de diferenciação e desenvolvimento celular, além de auxiliar na estabilidade genômica. A metilação geralmente ocorre em citosinas seguidas por guaninas, que estão concentrados em trechos ricos de CpG que são denominados de ilhas CpG, que frequentemente estão em regiões promotoras dos genes. A hipometilação leva a perda da heterocromatina que influencia na ativação de genes, e a hipermetilação ocasiona na compactação anormal da cromatina levando ao silenciamento gênico. Portanto, o processo tanto de metilação quanto de desmetilação do DNA são importantes para manter a integridade genômica e a expressão gênica e, quando desregulados, podem levar ao desenvolvimento tumoral (21,22).

Já as modificações pós-traducionais em histonas ocorrem na grande maioria em caudas N-terminais de histonas, e as alterações se classificam em acetilação, fosforilação, ubiquitinação, sumoilação e metilação (22), sendo essa última a alteração estudada no presente trabalho. A metilação em histonas ocorre em sua grande maioria em resíduos de lisina (K) catalisadas por enzimas chamadas de histonas metiltransferases (HMTs), que serão abordadas mais adiante, sendo responsáveis pela transferência de um, dois ou três grupos metil realizando a mono, di ou trimetilação, respectivamente, nos resíduos de lisina que tem como função ativar ou reprimir a expressão gênica (23). Isso ocorre por que as modificações nas histonas influenciam a acessibilidade da cromatina participando do controle da expressão gênica. Portanto, há uma relação entre a genética e a epigenética durante o desenvolvimento do câncer, e as alterações epigenéticas tem chamado atenção pela

sua natureza reversível que leva a oportunidades terapêuticas mais específicas (21,22).

#### 1.4. O câncer no Brasil

Estima-se que para o triênio 2023-2025 ocorrerá 704 mil casos novos de câncer no Brasil. Tirando os casos de câncer de pele não melanoma, a estimativa dos tumores mais incidentes na população brasileira durante esse período será de 74 mil em cânceres de mama (10,5%); 72 mil de próstata (10,2%); 46 mil de cólon e reto (6,5%); 32 mil de pulmão (4,6%); e 21 mil de estômago (3,1%) (24).

##### 1.4.1. Câncer de mama

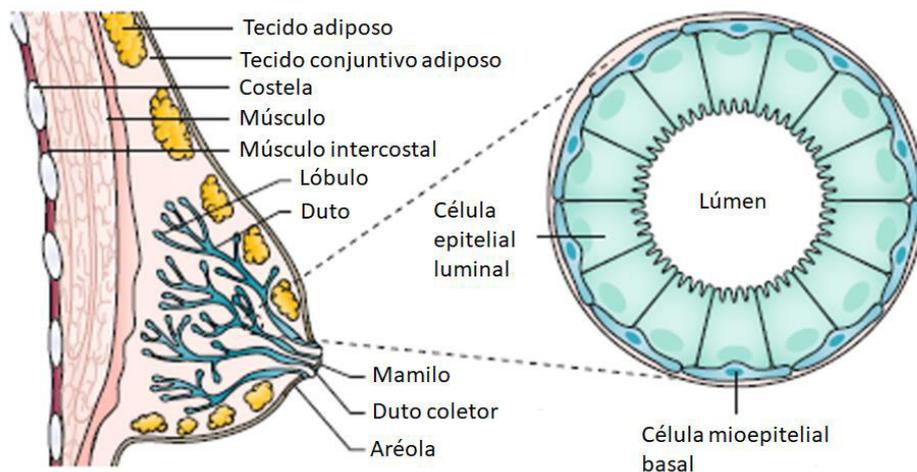
O câncer de mama no Brasil é o mais incidente entre a população do sexo feminino em todas as regiões do país, e o risco estimado durante o período de 2023 a 2025 será de 66,54 casos a cada 100 mil mulheres (24). Além da predisposição genética, outros fatores de risco associam-se a condições hormonais ou reprodutivas, como por exemplo a gravidez tardia e algumas causas relacionadas ao estilo de vida como obesidade, sedentarismo e consumo excessivo de álcool (24,25).

Na prática clínica, o diagnóstico é realizado através de análises histológicas e características moleculares (Figura 2) (25). O câncer de mama receptor hormonal positivo é caracterizado pela expressão de receptor de estrogênio (ER+) e/ou receptor de progesterona (PR+), e são divididos em dois subtipos: Luminal A e B.

O Luminal A é o subtipo mais comum abrangendo 50% dos casos de câncer de mama, caracterizado por ser ER+. Geralmente possui um melhor prognóstico quando comparado com o Luminal B, e tem uma melhor resposta a terapia endócrina. Histologicamente, são de baixo grau e usualmente não é do tipo invasivo (26). O Luminal B ocorre em cerca de 20% dos casos e é mais agressivo, além de possuir um pior prognóstico comparado com o Luminal A. É classificado em dois grupos: ER+/PR+, caracterizado por um alto índice de proliferação, e receptor tipo 2 do fator de crescimento epidérmico humano positivo (HER2+), caracterizado pela alta expressão de genes relacionados a proliferação (26). Pelo seu alto nível de proliferação celular, a quimioterapia costuma ser um bom candidato terapêutico (27).

O subtipo receptor tipo 2 do fator de crescimento epidérmico humano positivo (HER2+), abrange cerca de 20% dos casos diagnosticados e possui um mau prognóstico. HER2 é uma proteína que quando desregulada leva ao aumento de moléculas sinalizadoras de crescimento e de expressão de genes relacionados a proliferação celular, que contribui para um elevado crescimento tumoral resultando em uma evolução rápida da progressão do tumor. O câncer de mama HER2+ pode ser tanto positivo quanto negativo para os receptores hormonais. É invasivo com elevado risco de recorrência, que ocorre geralmente dentro de um período de 5 anos após o tratamento (26).

Por fim, o subtipo triplo negativo (TNBC) é caracterizado por ser negativo para ER, PR e HER2. É um tumor heterogêneo com fenótipo clínico distinto que compõe de 15% a 20% dos casos relatados. O TNBC possui padrões bastante distintos de metástase, podendo metastatizar para regiões distantes, como por exemplo para vísceras e cérebro, tendo mau prognóstico e um comportamento biológico bastante agressivo, além de possuir um crescimento mais acelerado. É mais comum em mulheres jovens antes da menopausa (26). O câncer de mama do tipo basal é bastante comum sendo tipicamente TNBC, de alto grau associado ao mau prognóstico de comportamento agressivo. É dividido em dois subgrupos, sendo o BL1 que é enriquecido em genes que estão envolvidos com a proliferação celular e a resposta aos danos do DNA, e o BL2, que é enriquecido em vias de sinalização para receptor de fator de crescimento (27).



**Figura 2:** Representação esquemática histológica da mama. (Adaptado de Harbeck N. et. al. Breast cancer. Nat Rev Dis Primers. 2019).

Dessa forma, percebe-se que o câncer de mama é altamente heterogêneo e que necessita de testes para saber se o tumor possui ER, PR, HER2 ou se é TNBC para um diagnóstico mais preciso, auxiliando dessa forma a escolha da melhor abordagem terapêutica.

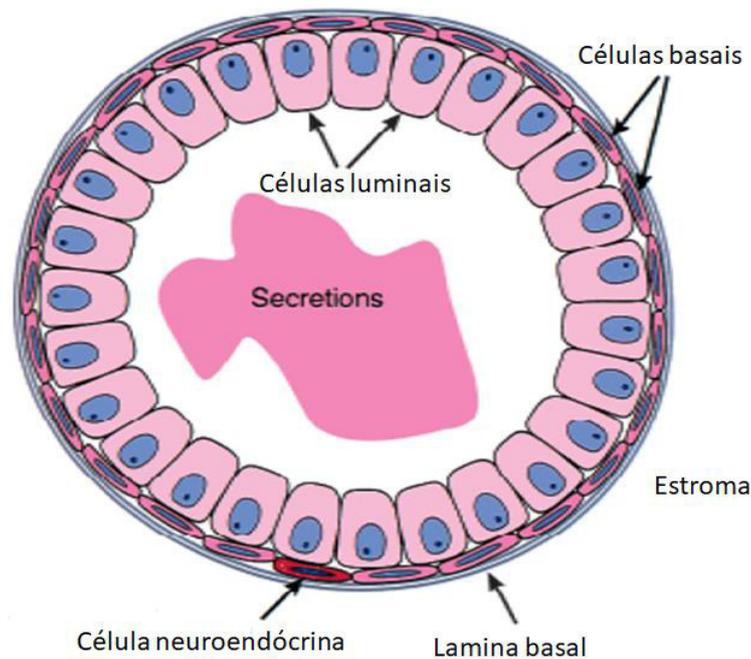
#### 1.4.2. Câncer de próstata

No Brasil, o câncer de próstata é o tumor mais incidentes dentre a população masculina, com estimativa de risco de 67,86 novos casos a cada 100 mil homens (24). O câncer de próstata surge através de um processo de vários passos, que resumidamente engloba a iniciação através da neoplasia intraepitelial prostática, que são lesões consideradas pré-neoplásicas, progredindo para o câncer de próstata localizado e, por fim, o desenvolvimento do adenocarcinoma de próstata avançado com invasão local, podendo resultar em metástase que geralmente ocorre em linfonodos próximos, seguido do fígado pulmão e ossos (28).

As células epiteliais que possuem alta relação com a carcinogênese prostática incluem as células luminais, que são as mais predominantes e dependentes de andrógenos, produzem proteínas secretoras prostáticas e são caracterizadas pela expressão de receptores de androgênio (AR); as células basais que também possuem AR porém em baixos níveis e diferentemente das luminais, são responsáveis pela expressão de fatores de proteção a danos ao DNA; por fim, as células neuroendócrinas são responsáveis por fornecer sinais parácrinos que auxiliam no suporte do crescimento das células luminais. Essas células são independentes de AR, e o seu acúmulo é caracterizada pela agressividade do câncer de próstata (Figura 3) (29).

É caracterizado por uma grande heterogeneidade por possuir uma justaposição de glândulas benignas, lesões pré-neoplásicas e neoplásicas, além de possuírem características de multifocalidade, que revelam uma distinção genética de células de origem não clonal (29). Para detecção, o exame de dosagem do antígeno específico prostático, conhecido por PSA, é utilizado na prática clínica por estar relacionado ao câncer de próstata quando encontrado em altos níveis. Porém, a biópsia tecidual continua sendo a melhor abordagem diagnóstica, já que o PSA pode ser encontrado em níveis elevados em alguns homens que não possuem câncer de próstata (30).

Abate-Shen and Shen



**Figura 3.** Esquema de células que compõem o ducto prostático. Secretions: Secreções. (Adaptado de Abate-Shen C, Shen MM. Molecular genetics of prostate cancer. Genes Dev. 2000).

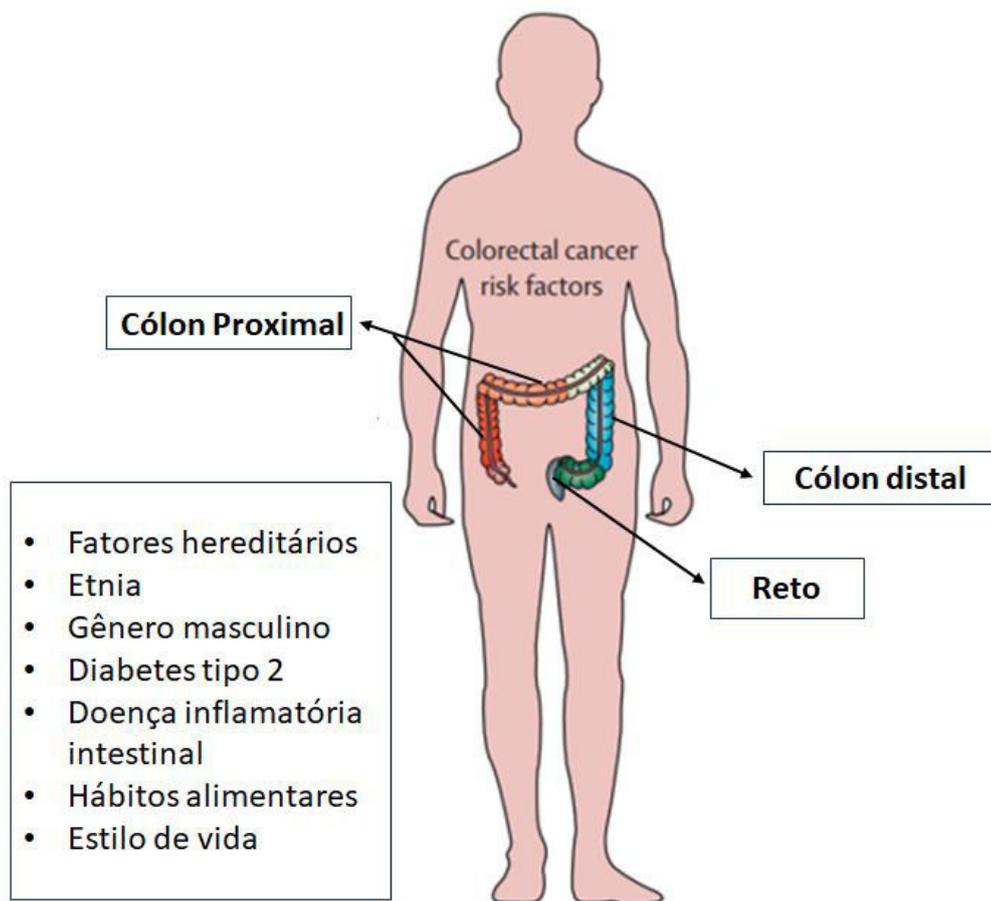
Os fatores de risco incluem a idade avançada, etnia, que abarcam fatores tanto socioeconômicos quanto biológicos, fatores genéticos e histórico familiar. Outros fatores englobam estilo de vida, como dieta com alta ingestão de carne processada e gordura saturada de origem animal, obesidade, consumo de álcool, cigarro e sedentarismo, e fatores ambientais como exposição a compostos químicos e radiação (30).

#### 1.4.3. Câncer colorretal

O câncer colorretal (CRC) ocupa a terceira posição entre os tumores mais incidentes no Brasil e, de modo geral, seu risco estimado para o triênio 2023-2025 é de 21,10 casos a cada 100 mil habitantes. Em relação ao gênero, o risco estimado para população masculina é de 20,78 a cada 100 mil homens e para população feminina é de 21,41 a cada 100 mil mulheres (24).

CRC é definido pela localização tumoral que vai especificar o seu subtipo, que pode ocorrer no cólon proximal, cólon distal e reto. Dessa forma, percebe-se que o CRC é amplamente heterogêneo e tanto o prognóstico quanto a resposta terapêutica

é influenciada pela localização tumoral, já que as características moleculares se diferem entre a região proximal, distal e retal (Figura 4) (31,32). A grande parte do CRC surge através da formação de um pólipó, que é uma lesão precursora neoplásica, e leva um período de 10 a 15 anos até se desenvolver em câncer. Outra forma conhecida que leva ao desenvolvimento tumoral é o CRC derivado de células tronco, que passa por um processo de acúmulo de alterações genéticas e epigenéticas que acarretam na inativação de genes supressores tumorais e a ativação de oncogenes (32).



**Figura 4.** Representação da região colorretal e descrição dos principais fatores de risco. (Adaptado de Dekker E. et.al. Colorectal cancer. Lancet. 2019).

O câncer colorretal é uma doença assintomática durante grande parte de seu desenvolvimento, e os sintomas geralmente só começam aparecer quando o paciente se encontra em estágio avançado. Alguns dos sintomas mais comuns englobam sangramento retal oculto nas fezes ou evidente, mudança de hábitos intestinais, anemia e fortes dores abdominais (32).

Em relação aos fatores de risco, a idade é levada em consideração pois a incidência é elevada em pessoas com mais de 50 anos. Os fatores genéticos se dividem em dois tipos, sendo classificados em casos esporádicos, do qual mutações somáticas e alterações epigenéticas são acumuladas ao decorrer da vida, e fatores hereditários. O estilo de vida também pode influenciar o desenvolvimento do tumor, já que o sedentarismo eleva o nível de gordura, inflamação e alto peso corporal, que influencia a carcinogênese por estar associado a inflamação de baixo grau, podendo promover a transcrição de genes que estão envolvidos com a metaplasia; dieta pobre, já que o alto consumo de carne vermelha e comidas processadas, carboidratos refinados, baixa ingestão de frutas, verduras, legumes e cálcio facilita a respostas inflamatórias; outro fator é a disbiose da microbiota intestinal, onde foi observado que infecção causada por *Fusobacterium nucleatum*, *Bacteroides fragilis* e *Streptococcus gallolyticus* podem elevar o risco de CRC (31–35).

#### 1.4.4. Câncer de pulmão

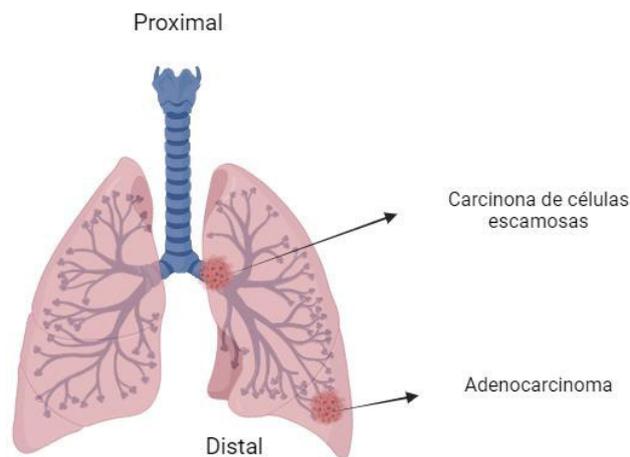
O câncer de pulmão ocupa a quarta posição entre os tumores mais incidentes na população brasileira, com o risco estimado para o triênio 2023-2025 é de 15,06 casos a cada 100 mil habitantes. Ao estratificar entre gêneros, o risco estimado é de 17,06 casos a cada 100 mil homens e 13,15 casos a cada 100 mil mulheres (24).

O câncer de pulmão se difere através de características histológicas, moleculares e etiológicas, sendo divididos em carcinoma de pulmão de células pequenas (SCLC) e carcinoma de pulmão de células não pequenas (NSCLC). Esse último se subdivide em carcinoma de pulmão de células escamosas (LUSC), adenocarcinoma de pulmão (LUAD) e carcinoma de células grandes (36). O presente estudo teve como foco os subtipos LUSC e o LUAD.

Acredita-se que o câncer de pulmão ocorre através de múltiplas mudanças patológicas na região da mucosa respiratória (36), sendo que o adenocarcinoma de pulmão surge geralmente nas vias aéreas distais, com histologia glandular e comumente expressam biomarcadores de origem pulmonar distal, como o Fator de transcrição da tireóide-1 (TTF-1) e queratina 7 (KRT-7). O carcinoma de pulmão de células escamosas comumente ocorre nas vias aéreas proximais, e associa-se fortemente ao tabagismo e a inflamação crônica de pulmão quando comparado com o LUAD (Figura 5). O LUSC é caracterizado pela diferenciação escamosa, que advém

do epitélio colunar pseudoestratificado, que é conhecido por reverter a região da traqueia e as vias aéreas superiores. (37).

Entre os fatores de risco, o tabagismo é o principal hábito associado ao desenvolvimento do câncer de pulmão, que aumenta o risco do câncer também em fumantes passivos. Isso se dá pelo fato de o cigarro conter mais de 4 mil substâncias químicas, com ao menos 50 dessas sendo classificadas como carcinogênicas. Outros fatores de risco incluem a idade, gênero e fatores socioeconômicos, além de fatores genéticos e histórico de câncer familiar. A exposição ocupacional a produtos químicos, como níquel, cromo, sílica entre outros, a exposição a poluição do ar devido a mudanças ambientais, queima de carvão e biomassa para cozinhar e/ou aquecer, e a alimentação, com alta ingestão de carne vermelha, gordura saturada e lipídeos, também aumentam o risco de câncer de pulmão (38).



**Figura 5.** Representação esquemática pulmonar evidenciando o carcinoma de pulmão de células escamosas e adenocarcinoma de pulmão. (Criado com BioRender.com).

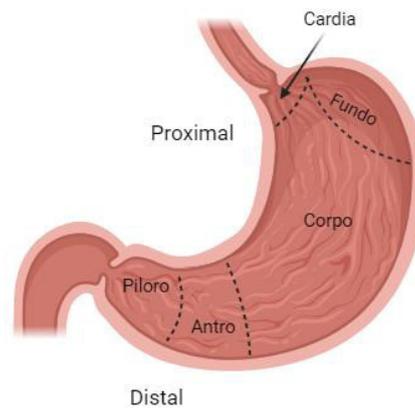
#### 1.4.5. Câncer de estômago

O câncer de estômago ocupa a quinta posição entre os tumores mais incidentes no Brasil, com o risco estimado em 9,94 casos novos a cada 100 mil habitantes para o triênio de 2023-2025. Avaliando apenas a população feminina, o risco estimado é de 7,36 casos a cada 100 mil mulheres e, para população masculina, 12,63 casos a cada 100 mil homens. Sua incidência é de quase duas vezes mais entre os homens do que em mulheres (24).

O câncer gástrico é separado anatomicamente em dois tipos: adenocarcinoma gástrico verdadeiro, conhecido como não cárdia, e o adenocarcinoma da transição esofagogástrica, também conhecido como câncer gástrico cárdia. A maioria dos tumores gástricos é do tipo adenocarcinoma, sendo amplamente heterogêneo por haver diferenças na arquitetura, diferenciação e crescimento celular, além das diferenças entre as características moleculares (39). Segundo a classificação de Lauren, pode-se classificar o câncer de estômago levando em consideração dois tipos histológicos, sendo o carcinoma difuso, que é pouco diferenciado e não há formação glandular. É localizado na região proximal, e possui um pior prognóstico por ter maior probabilidade de eventos metastáticos. O outro tipo é denominado de carcinoma do tipo intestinal, que vai de moderado a bem diferenciado havendo formação de estruturas glandulares. Geralmente ocorre na região distal do estômago e é caracterizado por ulcerações (Figura 6) (40).

Entre os fatores de risco, a infecção por *Helicobacter pylori* continua sendo a principal causa de câncer de estômago que afeta a região distal (39). Isso ocorre porque a maioria das cepas de *H. pylori* possui uma citotoxina associada ao gene A (CagA), que é uma oncoproteína responsável por afetar a expressão de outras proteínas envolvidas nas vias de sinalização celular (41). A infecção por *H. pylori* modifica processos celulares como a proliferação e apoptose, além de influenciar alterações epigenéticas em genes supressores tumorais que podem acarretar em inflamações associadas a oncogênese (39). Outros fatores de risco englobam infecção pelo vírus Epstein-Barr, idade avançada, baixo status socioeconômico, tabagismo, consumo excessivo de álcool, baixo consumo de frutas e vegetais, anemia perniciosa e fatores genéticos (41).

O câncer gástrico é assintomático, o que dificulta seu diagnóstico em fase inicial. Quando diagnosticado com os sintomas já presentes, provavelmente a doença se encontra em estágio avançado, geralmente incurável sendo recomendado apenas tratamento paliativo. Os sintomas mais comuns são dispepsia, anorexia ou sentimento de saciedade de forma precoce, perda de peso e dores abdominais (41).



**Figura 6.** Representação esquemática gástrica. Cardia, fundo e corpo representam a região proximal; Antro e piloro representam a região distal. (Criado com BioRender.com).

#### 1.4.6. Câncer de útero

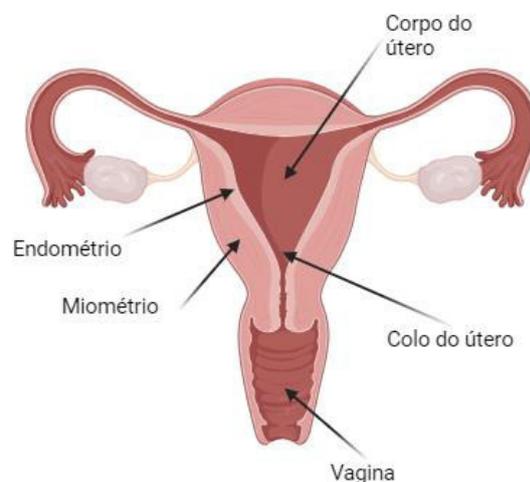
O câncer de útero é amplamente heterogêneo e possui diferenças em relação ao risco estimado e sua incidência, além dos fatores de risco, a depender do local afetado e do tipo do tumor que acomete o útero. O câncer de colo de útero possui um risco estimado de 15,38 novos casos a cada 100 mil mulheres, e ocupa a sexta posição entre os tumores mais incidentes quando avaliado de forma geral na população brasileira e, quando avaliado apenas entre as mulheres, é o terceiro mais incidente. Já o câncer do corpo uterino pode iniciar em diferentes partes do órgão e tem a região do endométrio como a mais acometida. Seu risco estimado é de 7,08 casos a cada 100 mil mulheres e ocupa o sétimo lugar entre os tumores mais incidentes na população feminina (Figura 7) (24).

Em relação ao câncer de colo de útero, a principal causa é a infecção por algumas cepas de HPV. A infecção é assintomática e por essa razão, pode levar um período de 10 a 15 anos até que ocorra uma mudança no colo do útero. Outros fatores que contribuem para o câncer de colo de útero é a idade precoce ao iniciar relações sexuais, tabagismo, que contribui com a destruição de células cervicais e comprometimento do sistema imune, deixando a paciente mais suscetível a infecções, e o uso prolongado de anticoncepcional oral por um período maior que 5 anos (42).

Já para o câncer do corpo do útero foi proposto dois tipos de classificação de acordo com a patologia, sendo: Tipo I, que é caracterizado pelo alto ou moderado grau de diferenciação, invasão moderada do miométrio com um prognóstico favorável; Tipo

II, que é pouco diferenciado com invasão profunda ao miométrio, com chances de processos metastáticos para pele e linfonodos, com prognóstico desfavorável. Em relação a características histopatológicas, é classificado em endometriode e suas variantes. Uma parcela pequena apresenta uma diferenciação mesenquimal, sendo então classificado como carcinosarcoma (43,44).

O carcinoma endometriode é o mais comum do câncer de endométrio e sua classificação entre o tipo usual e suas variantes vão depender da especificidade da diferenciação celular. É dividido ainda em 3 graus: Grau 1, que é composto em grande maioria por glândulas com células bem diferenciadas; Grau 2, que contém cerca de 6% a 50% de crescimento tumoral sólido; Grau 3, que possui mais de 50% de crescimento tumoral sólido. O carcinosarcoma, por sua vez, é um tumor uterino raro de natureza agressiva que geralmente projeta-se através do colo do útero e possui um mau prognóstico. Possui características de carcinoma e sarcoma e, por isso, é considerado um tumor bifásico. Seu processo metastático é similar ao que ocorre no carcinoma endometrial e é composto majoritariamente por componentes carcinomatosos (43,44).



**Figura 7.** Representação esquemática uterina. (Criado com BioRender.com).

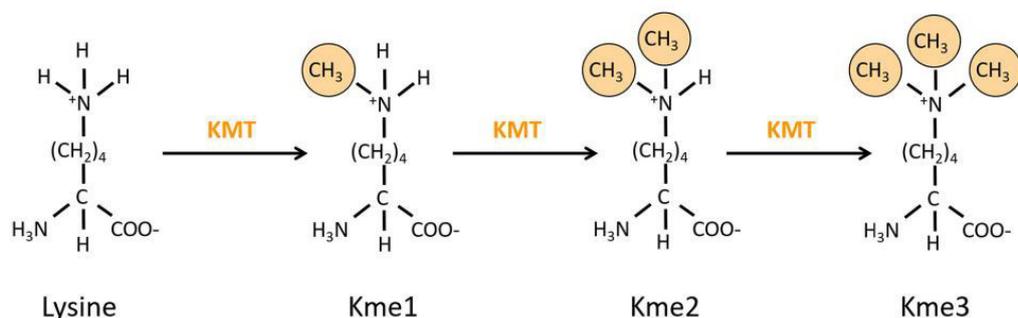
Entre os fatores de risco, os fatores metabólicos contribuem para o desenvolvimento do câncer endometrial através da obesidade, diabetes, hipertensão e síndrome do ovário policístico. Outras questões como fatores reprodutivos que incluem a nuliparidade, infertilidade, menarca precoce e menopausa tardia, além de terapia hormonal na menopausa, fatores genéticos e histórico familiar também contribuem para o risco de câncer do corpo do útero (45).

### 1.5. Metiltransferases e o câncer

As histonas metiltransferases (HMTs) são proteínas responsáveis por alterações reversíveis pós-traducionais, que ocorrem em resíduos de lisina e arginina em caudas N-terminais de histonas. O processo de metilação realizado por lisinas metiltransferases (KMT) nos resíduos de lisina divide-se em duas classes de enzimas de acordo com seu domínio catalítico: uma que possui o domínio conservado chamado SET (Supressor of variegation, Enhancer of zester, e Irithorax), composto por 130 aminoácidos e sendo essa a maior parte das KMT descritas, e o DOT1L, que não possui o domínio SET. Ambas as classes, porém, utilizam S-adenosil-L-metionina (SAM) como doador de grupo metil (46,47).

Portanto, um dos mecanismos responsáveis pela regulação da cromatina envolve as modificações de histonas, que ocorrem através da metilação em resíduos de lisina que, por sua vez, podem receber até três grupos metil, denominados de mono, di e trimetilação (Km1, Km2 e Km3, respectivamente) (Figura 8). A depender do grau de metilação e a posição da lisina que recebe a alteração é que será determinado a função e a consequência em relação aos estados da cromatina e, dessa forma, será determinado a acessibilidade ao DNA (47–49). Em humanos, os locais metilados melhores caracterizados são as marcas encontradas na histona 3 lisinas 4, 9, 27, 36 e 79, e na histona 4 lisina 20 (48).

Os processos de metilação estão envolvidos na regulação da expressão gênica, já que suas marcas podem servir tanto para repressão quanto para ativação gênica atuando, dessa forma, na regulação das informações genéticas e as funções biológicas como, por exemplo, a transcrição, replicação e manutenção dos cromossomos (47,49).



**Figura 8:** Representação de mono, di e trimetilação em resíduos de lisina. (Yang C. et.al. Histone methyltransferase and drug resistance in cancers. J Exp Clin Cancer Res. 2020).

De modo geral, inicialmente as proteínas contendo o domínio SET se encaixam em dois complexos proteicos, sendo o complexo Polycomb (PcG), associados a heterocromatina e conseqüentemente a repressão gênica, e o Trithorax (TrxG), relacionados a ativação gênica (50). O E(z) faz parte do PcG e em associação com o Su(var)3-9 são responsáveis pela adição de marcas heterocromáticas enquanto removem as marcas eucromáticas (51). O PcG é composto por proteínas que reprimem a transcrição e auxiliam a manutenção da diferenciação celular mantendo a identidade celular. No processo carcinogênico, a tumorigênese é induzida através da repressão da transcrição de genes supressores tumorais (46).

O TrxG, por sua vez, é um grupo mais heterogêneo quando comparado ao PcG, e estão envolvidos em processos de ativação transcricional. Isso ocorre pelo fato do TrxG estar associado a indução da trimetilação na lisina 4 da histona 3 (H3K4me3), marca essa que está amplamente presente na região promotora de genes ativamente transcritos (52,53).

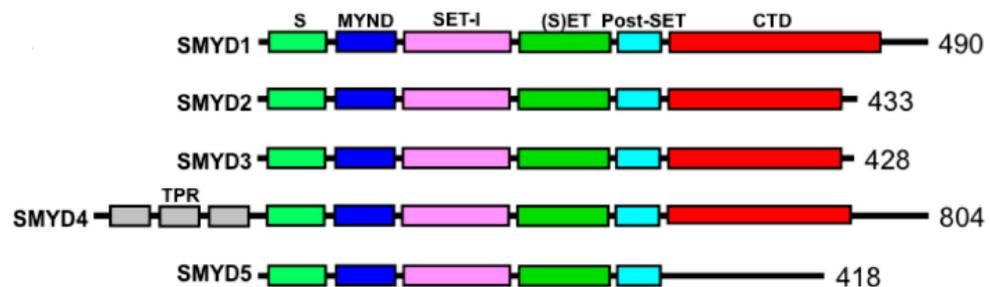
Portanto, o processo evolutivo do câncer também se dá através de alterações epigenéticas que ocorrem quando há alguma disfunção no funcionamento das HMTs, o que pode afetar as características biológicas de células tumorais. As proteínas do complexo PcG/TrxG mostram-se então essenciais para a manutenção da expressão gênica, possuindo funções importantes em diversos processos celulares já que as mesmas agem tanto como supressores tumorais quanto oncogenes, a depender do contexto tumoral (46,49,52).

#### 1.6. Família SMYD de metiltransferases

A família SMYD é composta por cinco membros (SMYD1-5) que possuem o domínio SET, responsável pela catálise da metilação em resíduos de lisinas e que é “dividido” pelo domínio MYND (Myeloide - Nervy – DEAF1), que é responsável por facilitar a interação proteína-proteína por conter o motivo dedo de zinco, que participa da região N-terminal e contribui para a interação com o domínio SET. O domínio MYND, portanto, é o que diferencia a família SMYD das outras lisinas metiltransferases (Figura 9). Com a divisão do domínio SET, o “S” auxilia na ligação a cofatores ou na interação proteína-proteína juntamente com o domínio MYND, além

de conter a função catalítica da enzima. As regiões pós-SET e SET-I auxiliam nos processos de ligação à cofatores e substratos, ou a estabilidade da estrutura proteica. Os membros da família SMYD podem metilar proteínas tanto histonas quanto não histonas (54,55).

Além dos domínios citados acima, exceto pelo SMYD5, o domínio C-terminal (CTD) também se encontra nessa família de genes que, estruturalmente, se assemelha ao tetratricopeptídeo (TRP), importante para o processo de ligações a co-chaperonas. No SMYD4, esses dois domínios são maiores, que faz com que o SMYD4 tenha quase que o dobro do tamanho dos outros membros da família SMYD (55).



**Figura 9:** Representação de domínios encontrados entre as proteínas da família SMYD de metiltransferases. S, SET-I (S)ET e Post-SET representam a totalidade do domínio SET; MYND (Myeloide - Nervy – DEAF1) é o domínio responsável pela interação proteína-proteína; CTD t o domínio C-terminal; TRP é o domínio tetratricopeptídeo. (Spellmon N. et.al. Structure and function of SET and MYND domain-containing proteins. Int J Mol Sci. 2015).

A família SMYD, portanto, é capaz de metilar resíduos específicos de lisina e, com isso, podem tanto silenciar quanto potencializar a expressão de genes alvos. Contudo, sua disfunção pode acarretar a indução da oncogênese através de modificações epigenéticas e, de modo singular, cada um dos cinco membros da família SMYD contém variações de acordo com a distribuição entre os tecidos e suas funções (56).

### 1.6.1. SMYD1

Inicialmente foi identificado como uma proteína específica do músculo cardíaco e esquelético, participando do processo de diferenciação celular e morfogênese durante o desenvolvimento embrionário (57). Foi observado também que a ausência do *SMYD1* acarreta em defeitos ao ventrículo direito por afetar a diferenciação de cardiomiócitos (54). O *SMYD1* é o único dentre a família que possui uma alta

expressão de suas várias isoformas em tecidos cardíacos e esqueléticos, além de ser necessário para o controle do crescimento cardíaco (56).

Foi sugerido que sua atividade catalítica seja controlada pelo domínio C-terminal, que pode adquirir duas formas diferentes a depender da condição ambiental ao ligar-se a substratos alvos. No processo de metilação, o *SMYD1* é responsável pela marca no resíduo de lisina 4 da histona 3 (H3K4), um resíduo conhecido por fazer parte da iniciação da transcrição (54,56).

No estudo da carcinogênese, o *SMYD1* ainda é pouco elucidado, porém já foi descrito que sua desregulação teve relação com o rabdomyosarcoma (58). No câncer de mama, foi observado que houve uma redução de sua expressão a nível de mRNA (59). No linfoma esplênico da zona marginal, foram encontradas mutações somáticas (60). Alterações na expressão do mRNA também foram observadas em tumores do sistema digestivo (61). E, por fim, no câncer de próstata foi observado que houve alteração no CNA que associava-se a sobrevida livre de progressão (62).

#### 1.6.2. *SMYD2*

O *SMYD2* se encontra em níveis elevados nos tecidos do coração, cérebro, fígado, rim, timo e ovário, além de ser encontrado também em tecido embrionário referente ao coração e hipotálamo, sugerindo que o *SMYD2* seja necessário para o desenvolvimento (63).

Foi observado que o local de ação do *SMYD2* ocorre na lisina 4 da histona 3 quando em conjunto com a HSP90 (heat shock protein), marca essa responsável pela ativação da expressão gênica, enquanto que o *SMYD2* sozinho, por ter um enfraquecimento em sua atividade, é responsável pela metilação na lisina 36 da histona 3, marca essa responsável pela repressão da expressão gênica. Além do mais, foi observado que a elevação da expressão do *SMYD2* pode ocasionar a *upregulação* de trinta e sete genes que participam da remodelagem da cromatina, ciclo celular e regulação da transcrição (64).

O *SMYD2* também é conhecido por metilar proteínas não histonas, como por exemplo a p53 e Retinoblastoma (Rb). Na p53, o *SMYD2* é responsável pela monometilação da K370, que leva a redução da eficiência do p53 de se ligar ao DNA por reprimir sua atividade. Dessa forma, o *SMYD2* tem ação oncogênica por ocasionar

a eliminação da função do supressor tumoral p53. Já no Rb, o SMYD2 é responsável por metilar a K860, que leva a condensação da cromatina e por conseguinte, leva a repressão da expressão gênica (65). Outro supressor tumoral que o *SMYD2* auxilia na regulação é o PTEN. A metilação do PTEN na K313 pelo SMYD2 ocasiona em sua redução de supressor tumoral, promovendo a proliferação das células cancerígenas (66).

Portanto, o *SMYD2* é um dos membros da família SMYD melhor descrito no âmbito da carcinogênese, sendo reconhecido por ser um regulador de supressores tumorais, e sua superexpressão está associada ao desenvolvimento de diversos tumores, como o pancreático, adenocarcinoma pulmonar, o câncer de mama triplo negativo, o câncer cervical, entre outros (67–70).

### 1.6.3. *SMYD3*

Semelhante ao *SMYD1*, o *SMYD3* também possui atividade durante o processo de diferenciação do músculo esquelético por conter um nível elevado de sua expressão em mioblastos (71). Foi observado que o SMYD3 em conjunto com HSP90A realizam a metilação no resíduo de lisina 4 da histona 3 (H3K4), sendo responsável pela di- e tri-metilação. Observou-se também que o SMYD3 pode agir na histona 4 nos resíduos de lisina 5 e 20 (H4K5 e H4K20, respectivamente). A marca H4K5 mostrou-se necessária para manter a alta expressão do *SMYD3* em células cancerígenas (72–74).

Em relação a metilação em proteínas não histonas no âmbito tumoral, o SMYD3 é responsável por metilar a K831 do receptor de tirosina quinase VEGFR1, conhecido pela sua função na angiogênese, associando-se ao crescimento e progressão tumoral e, portanto, o aumento do *SMYD3* pode levar a invasão e a metástase de células cancerígenas (75). Também é responsável pela metilação na K260 da MAP3K2, que é um evento importante para a regulação da via de sinalização Ras em células cancerígenas. Nesse contexto, o *SMYD3* auxilia a promoção do Ras durante o desenvolvimento e a progressão de células cancerígenas, já que essa via é responsável pela estimulação da diferenciação e multiplicação celular (76).

Dentre a família *SMYD*, o *SMYD3* é um dos membros que mais possui estudos que relaciona seu envolvimento durante os processos carcinogênicos. Isso porque já é descrito pela literatura a função que o *SMYD3* exerce durante a progressão e

invasão do câncer através da regulação transcricional (72), associando-se a um mau prognóstico quando superexpresso por levar a proliferação, invasão e metástase tumoral. Sabe-se também que o *SMYD3* é capaz de reprimir a expressão do p53 e promover a transição epitélio-mesenquimal em células tumorais (71).

Portanto, a alteração da expressão do *SMYD3* já foi descrita em diversos tipos tumorais, como por exemplo no câncer colorretal e hepatocelular, no câncer de mama, câncer de pulmão de células não pequenas, câncer pancreático, leucemia linfóide crônica, entre outros (59,72,77–79).

#### 1.6.4. *SMYD4*

Em relação a saúde humana, há poucas informações disponíveis sobre o *SMYD4*. Porém, um estudo realizado em peixe zebra mostrou que seu homólogo *smyd4* possui atividade no desenvolvimento cardíaco. Foi sugerido também que seu local de ação ocorre na histona 3 lisina 4, realizando tanto a di- quanto a tri-metilação (80). Dessa forma, é importante ressaltar que a marca H3K4me3 é descrita como uma assinatura epigenética responsável pela ativação de supressores tumorais em células normais (81). Portanto, foi sugerido que o *smyd4* age como um regulador epigenético por conter função de metiltransferase e regulador negativo da *hdac1*, homólogo da HDAC 1 e 2 presentes em mamíferos, por impactar a função da mesma quando há um declínio do *smyd4* (80).

No ramo da oncologia, o *SMYD4* foi inicialmente descrito como supressor tumoral por reprimir a expressão do *Pdgfr- $\alpha$* , um receptor celular responsável pela ativação de vias relacionadas a proliferação e sobrevivência tumoral e, portanto, o silenciamento do *SMYD4* está associado a tumorigênese em células mamárias normais (82). Posteriormente, outros estudos corroboraram a hipótese do *SMYD4* agir como potencial supressor tumoral no câncer de mama por ter sido observado que o miR-1307-3p estimulava a proliferação celular ao inibir a expressão do *SMYD4* (83), e ter sido observado que o aumento de sua expressão associava-se a uma boa sobrevida livre de recidiva, sendo sugerido pelo autor que o *SMYD4* poderia servir como marcador de bom prognóstico para o câncer de mama (59).

Por outro lado, o *SMYD4* também foi associado a função oncogênica em tumores de células-tronco cancerígenas, que quando superexpresso, potencializa a expressão do *Nanog*, um fator transcricional presente em células-tronco que é uma

assinatura molecular frequente em células-tronco cancerígenas (84). Alterações no *SMYD4* foram descritas também em glioblastomas e em tumores do sistema digestivo, como por exemplo no carcinoma hepatocelular, adenocarcinoma gástrico e no carcinoma esofágico (61,85).

#### 1.6.5. *SMYD5*

Foi observado que o *SMYD5* participa da catálise da tri-metilação na lisina 20 da histona 4 (H4K20me<sub>3</sub>), e na mono-metilação da histona 3, lisina 36 e 37 (H3K36me<sub>1</sub> e H3K37me<sub>1</sub>, respectivamente) (86,87). A marca H4K20me<sub>3</sub> auxilia na formação da heterocromatina e conseqüentemente na repressão gênica por facilitar a ligação da Proteína da heterocromatina 1 (HP1). Com isso, foi visto que a redução do *SMYD5* afeta a capacidade da auto renovação e diferenciação de células tronco embrionárias por levar a redução de alguns reguladores de pluripotência, como por exemplo o Nanog, por reduzir o silenciamento de genes responsáveis pela diferenciação celular (86).

O *SMYD5* também participa do controle de respostas pró-inflamatórias através da metilação na H4K20 mediado pelo complexo NCoR/*SMYD5*, que é um complexo co-repressor que reprime a resposta inflamatória por restringir o receptor toll-like 4 (TLR-4) auxiliando, dessa forma, na manutenção da inflamação (88). Na hematopoiese, foi visto que há uma interação do *SMYD5* com a Nucleofosmina 1 (NPM1), uma proteína que regula a mielopoiese e, portanto, agem na regulação de células tronco hematopoiéticas (89).

Na carcinogênese, foi sugerido que o silenciamento do *SMYD5* esteja relacionado a instabilidade genômica, tumorigênese e na progressão tumoral por reduzir as marcas repressivas H4K20me<sub>3</sub> e H3K9me<sub>3</sub> em células cancerígenas (90). Contudo, apesar do *SMYD5* ser pouco estudado na área da oncologia, existem estudos promissores que mostram o envolvimento do mesmo no adenocarcinoma gástrico, no processo metastático do câncer de mama na colonização celular do pulmão, na leucemia mieloide aguda, no carcinoma hepatocelular, que ocasiona no aumento da proliferação, migração e invasão celular, e em leucemia linfóide aguda (89,91–94). Foi observado que o *SMYD5* pode agir tanto como supressor tumoral quanto oncogene, a depender do contexto tumoral (59,90).

## 2. OBJETIVOS

### 2.1. Objetivos gerais

Determinar o perfil de alterações moleculares nos genes da família *SMYD* e avaliar o seu valor prognóstico entre os tumores sólidos mais incidentes e na leucemia linfóide aguda (LLA).

### 2.2. Objetivos específicos

Foram divididos em duas etapas, sendo que para este primeiro capítulo os objetivos foram:

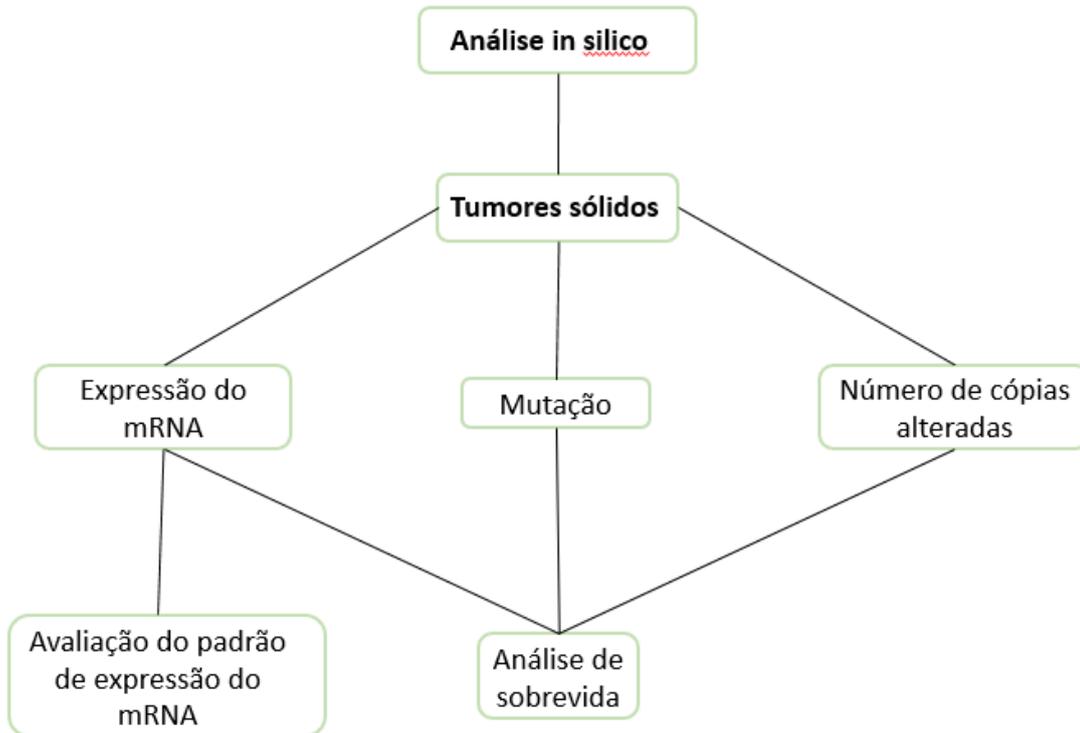
- Análise *in silico* em bases de dados genômicas para avaliação dos tipos de mutação, número de cópias alteradas e expressão aberrante do mRNA referente a todos os cinco membros da família *SMYD* entre as neoplasias malignas sólidas mais frequentes na população brasileira;
- Determinar as principais alterações câncer-específicas da família *SMYD* com intuito de propor novos biomarcadores moleculares.

Para o segundo capítulo, os objetivos específicos foram:

- Análise *in silico* em bases de dados genômicas para avaliar o perfil do número de cópias alteradas e expressão do mRNA da família *SMYD* na leucemia linfóide aguda.
- Iniciar a validação funcional de um dos alvos hiperexpresso em LLA através do nocaute gênico utilizando a tecnologia Crispr-Cas9.

### 3. METODOLOGIA

#### 3.1. Delineamento experimental



**Figura 10:** Delineamento experimental referente ao capítulo 1.

#### 3.2. Análise *in silico*

Os dados de mutação, número de cópias alteradas (CNA) e expressão do mRNA foram obtidos pela plataforma online “*cBio Cancer Genomics Portal*” ([www.cbioportal.org](http://www.cbioportal.org)). Essa plataforma reúne informações de alterações moleculares encontradas em pacientes oncológicos obtidos através de diversos estudos publicados, que possibilita avaliar um grande n amostral (Tabela 1). Para obtenção dos dados, os estudos e os genes de interesse são selecionados e os arquivos com as informações das alterações citadas acima são adquiridos na opção “*Download*”.

As abreviações utilizadas nesse estudo para representar os tumores foram retiradas da lista “*TCGA Study Abbreviations*”, encontrada na plataforma “*National Cancer Institute*” (<https://gdc.cancer.gov/>), sendo: Carcinoma Colorretal (CRC); Adenocarcinoma de Estômago (STAD); Câncer de Mama Invasivo (BRCA); Adenocarcinoma de Próstata (PRAD); Adenocarcinoma Pulmonar (LUAD); Carcinoma

de pulmão de células escamosas (LUSC); Carcinoma Endometrial do Corpo Uterino (UCEC); Carcinossarcoma Uterino (UCS).

**Tabela 1:** número de estudos selecionados e número de pacientes obtidos para avaliação das alterações moleculares referente a cada tumor analisado.

Tumores	Nº de estudos	Nº de pacientes
CRC	7	3.015
STAD	6	895
BRCA	11	4.398
PRAD	16	5.995
LUAD	11	3.595
LUSC	2	591
UCEC	3	818
UCS	2	79

Para avaliar as mutações, o arquivo “Mutations” foi obtido e as mutações foram separadas de acordo com o domínio do qual estava presente. Em seguida, as mutações foram agrupadas de acordo com as repetições entre pacientes com o mesmo tipo tumoral e entre os tumores avaliados, com intuito de analisar mutações frequentes, caso houvesse. Para validação de mutações que se apresentaram mais de uma vez em diferentes pacientes no mesmo tipo tumoral e/ou em diferentes tumores, a plataforma COSMIC (Catalogue Of Somatic Mutations In Cancer, em inglês) foi utilizada (95). Representações gráficas foram geradas pelo GraphPad Prism 8 e pelo software R.

A coorte com dados para mutação totalizou 67 pacientes com CRC, 92 pacientes com STAD, 834 pacientes com BRCA, 182 pacientes com PRAD, 128 pacientes com LUAD, 47 pacientes com LUSC, 62 pacientes com UCEC e 9 pacientes com UCS.

Para avaliar a frequência da alteração do número de cópias alteradas dos genes, o arquivo denominado “Copy-number Alterations” foi obtido. Esses dados utilizam o algoritmo GISTIC (Genomic Identification of Significant Targets in Cancer, em inglês), que indica o número de cópia de cada membro da família *SMYD* de acordo com o tumor avaliado, sendo: -2 indica perda homocigótica, -1 perda heterocigótica, 1

indica ganho de baixo nível e 2 indica amplificação. Os dados foram processados da seguinte forma: os pacientes foram agrupados de acordo com o nível da alteração do número de cópias do DNA que possuía, e em seguida foi calculado a frequência absoluta e depois a frequência relativa.

A coorte com dados para o número de cópias alteradas totalizou no CRC 616 pacientes para o *SMYD1*, *SMYD2*, *SMYD4* e *SMYD5*, e 1.943 para o *SMYD3*. 589 pacientes para STAD e 3.469 pacientes para o BRCA. Para o PRAD houve variação no n amostral, sendo 2.848 pacientes para o *SMYD1*, *SMYD2* e *SMYD5*, 3.890 para o *SMYD3* e 1.835 para o *SMYD4*. Para o LUAD também houve variação no n amostral sendo 1.109 pacientes para o *SMYD1*, *SMYD2* e *SMYD4*, 1.973 pacientes para o *SMYD3* e 1.158 pacientes para o *SMYD5*. 501 pacientes com LUSC, 620 pacientes com UCEC e 56 pacientes com UCS foram avaliados.

Para a análise de expressão do mRNA, o arquivo denominado “mRNA expression z-scores relative to diploid samples” foi obtido. (62,96). Para avaliar o nível de expressão, os dados foram plotados em diagrama de caixa (boxplot, em inglês) e o valor da mediana obtido foi usado para analisar os intervalos interquartis. Os valores acima da mediana são classificados como alta expressão, e os valores abaixo da mediana são classificados como baixa expressão.

A coorte com dados para expressão do mRNA totalizou 382 pacientes para CRC, 415 pacientes para STAD, 1.100 para BRCA, para o PRAD houve uma variação entre os genes, sendo 498 pacientes para o *SMYD1*, *SMYD4* e *SMYD5*, 486 pacientes para o *SMYD2* e 493 pacientes para o *SMYD3*, 169 pacientes para LUAD, 501 pacientes para LUSC, 177 pacientes para UCEC e 57 pacientes para UCS.

### 3.3. Mapa de calor

Os dados em Z-score relacionados à expressão de mRNA dos *SMYDs* nos tipos de câncer avaliados foram visualizados usando gráficos de mapa de calor (heatmap, em inglês) gerados por meio da ferramenta de análise e visualização de dados Flaski (<https://flaski.age.mpg.de>). Os dados foram dimensionados numericamente dentro do intervalo de  $\pm 2$ . O método de ligação de Ward com distância euclidiana foi empregado para agrupamento.

### 3.4. Análise de sobrevida

O banco de dados “Kaplan-Meier plotter” (<http://kmplot.com/analysis/>) foi usado para explorar a importância prognóstica das alterações nos cinco genes da família *SMYD*. Foram realizadas análises de sobrevida para determinar a correlação entre as alterações dos *SMYDs* e o prognóstico do tumor. Na análise para mutação, a opção “DNA pan cancer” foi selecionada e os pacientes foram categorizados em dois grupos: aqueles com mutações e aqueles com genes do tipo selvagem. Para a análise do nível de expressão de mRNA, a opção “mRNA pan cancer” foi selecionada e os pacientes foram divididos com base nos níveis de alta expressão e baixa expressão de acordo com a mediana.

A análise de sobrevida em relação à alteração do número de cópias foi realizada usando a plataforma *cBio Cancer Genomics Portal*, utilizando a mesma coorte dos estudos citados acima, com os pacientes categorizados em "grupo alterado", que continha pacientes com amplificação ou perda heterozigótica, e "grupo inalterado", que continha pacientes com CNA a níveis normais nos genes avaliados. O valor de  $P < 0,05$  foi considerado estatisticamente significativo.

### 3.5. Estatística

O programa estatístico *GraphPad Prism 8* foi utilizado para realização das análises estatísticas. Para a correlação do mRNA entre os genes da família *SMYD*, e a correlação entre CNA e mRNA no câncer de mama invasivo, foi realizado o teste de correlação de Pearson. O valor de  $p < 0,05$  foi considerado como estatisticamente significativo.

## 4. RESULTADOS

### 4.1. Perfil mutacional

Inicialmente, as mutações foram analisadas em cada um dos genes da família *SMYD* em todas as coortes disponíveis por meio do portal *cBio Cancer Genomics* para os oito tipos de tumores avaliados neste estudo. As mutações entre os cinco membros da família *SMYD* foram analisadas em um total de 9.696 pacientes de todas as coortes examinadas. Para a análise do *SMYD3*, um total de 11.804 pacientes estavam disponíveis para análise.

As mutações missense foram as mais prevalentes entre as mutações identificadas. O *SMYD1* apresentou a maior frequência de mutações missense, com 78 mutações distintas identificadas, seguido pelo *SMYD3* (46), *SMYD5* (44), *SMYD4* (39) e *SMYD2* (33). As mutações nonsense foram o segundo tipo de mutação mais comum, com seis casos detectados no *SMYD3*, quatro no *SMYD2*, três no *SMYD1* e *SMYD5* e duas no *SMYD4*. Outros tipos de mutação foram observados em menor número (Tabela 1).

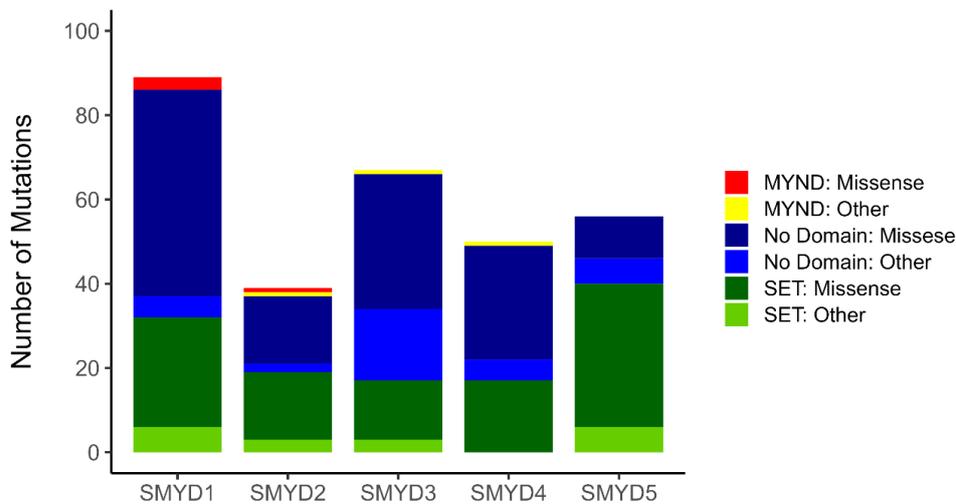
**Tabela 2:** Total de diferentes tipos de mutações somáticas encontradas nos membros da família *SMYD* em oito tumores sólidos analisados

Mutação	<i>SMYD1</i>	<i>SMYD2</i>	<i>SMYD3</i>	<i>SMYD4</i>	<i>SMYD5</i>
Missense	78	33	46	39	44
Nonsense	3	4	6	2	3
Splice	3	2	9	0	4
FS ins	1	2	1	0	0
FS del	2	0	5	3	3
IF del	0	1	0	0	2
IF ins	0	1	0	0	0
Nonstop	0	0	0	2	0
Fusion	0	0	0	2	0
Del	1	0	0	0	0
Ins	0	0	0	2	0

**Legenda:** (Ins = inserção; Del = deleção; IF ins = inserção in frame; IF del = deleção in frame; FS del = deleção frameshift; FS ins = inserção frameshift).

Posteriormente, o número total de mutações por domínio foi avaliado nos cinco genes. A análise foi dividida entre as regiões que codificam os domínios conservados SET e MYND e em regiões com domínios inespecíficos. As mutações missense foram avaliadas individualmente, enquanto que os demais tipos de mutação foram agrupados e analisados coletivamente. A maioria das mutações missense foram encontradas em domínios inespecíficos, totalizando 129 mutações, sendo o *SMYD1* responsável por 49 dessas mutações. Quanto aos domínios conservados, o domínio SET exibiu a maior frequência de mutação, com 107 mutações identificadas. Notavelmente, o *SMYD5* apresentou o maior número de mutações dentro do domínio SET, com 34 mutações documentadas (Figura 10)

Para obter um melhor panorama do perfil mutacional, as frequências de mutações foram avaliadas separadamente para cada gene em diferentes tipos de tumor. Observou-se que as frequências de mutação dentro da família *SMYD* eram relativamente baixas, com certos genes exibindo frequências superiores a 2%. Especificamente, o *SMYD5* teve uma frequência de mutação de 2,3% no câncer colorretal (CRC), enquanto o *SMYD1* mostrou uma frequência de mutação de 2,3% tanto no adenocarcinoma de pulmão (LUAD) quanto no carcinoma de pulmão de células escamosas (LUSC). No carcinoma endometrial (UCEC), as frequências de mutação foram de 3,3% para *SMYD1*, 2,1% para *SMYD2*, 2,1% para *SMYD3*, 2,1% para *SMYD4* e 2,4% para *SMYD5*, respectivamente. Além disso, no carcinossarcoma uterino (UCS), as frequências de mutação foram de 5,1% para *SMYD1* e 2,5% para *SMYD4* (Tabela 2). Estes resultados destacam que os tumores uterinos apresentaram as maiores frequências de mutação entre os tumores avaliados, com particular ênfase na frequência obtida pelo *SMYD1* no UCS.



**Figura 11.** Total de mutações somáticas separadas por domínios conservados e não conservados. O domínio SET é representado em verde escuro para mutação missense e verde claro para outras mutações; Domínio MYND é representado em vermelho para mutação missense e amarelo para outras mutações; Domínios inespecíficos representado em azul escuro para mutação missense e azul claro para outras mutações.

Nenhuma mutação específica entre os genes *SMYD* foi observada com frequência repetida entre pacientes com o mesmo tipo de tumor, em tumores diferentes ou em ambos.

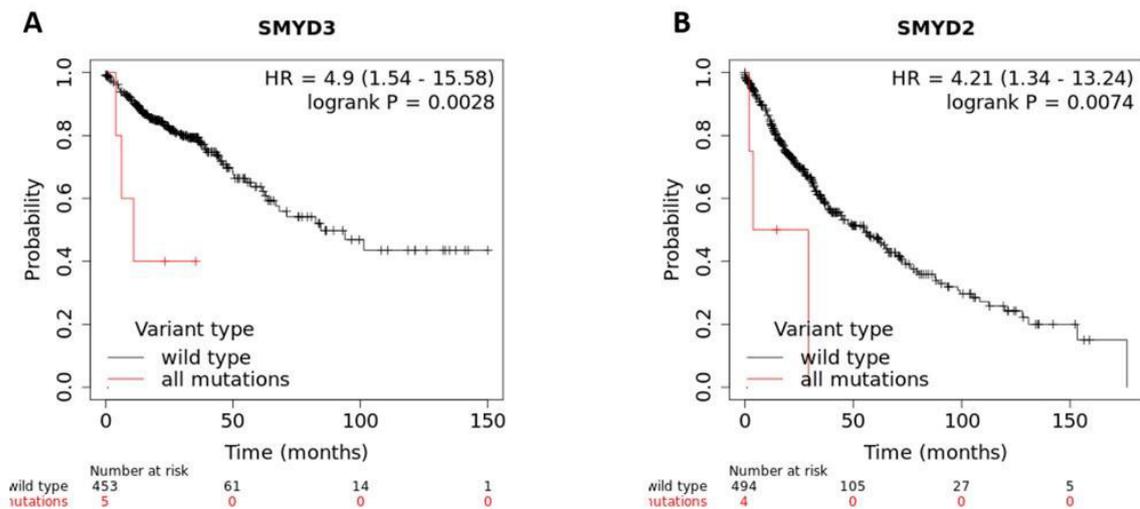
Em seguida, a plataforma Kaplan-Meier plotter foi utilizada para avaliar a correlação em potencial entre mutações em cada gene da família *SMYD* e o tempo

de sobrevida (OS) dentro de cada tipo de tumor examinado. Essa análise abrangeu a opção “all mutations”, que carece de especificação quanto aos tipos de mutação. Esta análise revelou que apenas mutações no *SMYD3* no carcinoma de cólon e *SMYD2* no LUSC tiveram um impacto significativo na sobrevida, levando a um prognóstico menos favorável em comparação aos casos do tipo selvagem. Porém, mais estudos são necessários para investigar sua relação com a progressão do câncer (Figura 11 A e B)

**Tabela 3.** Frequência de mutação em cada membro da família *SMYD* entre os oito tipos de tumores.

<b>Tumores</b>	<i>SMYD1</i>	<i>SMYD2</i>	<i>SMYD3</i>	<i>SMYD4</i>	<i>SMYD5</i>
CRC	1,4%	0,9%	1,0%	0,5%	2,3%
STAD	1,2%	0,9%	1,9%	0,8%	1,1%
BRCA	0,3%	0,2%	0,4%	0,4%	0,3%
PRAD	0,2%	0,1%	0,2%	0,1%	0,1%
<b>Câncer de pulmão</b>					
LUAD	2,3%	0,8%	0,7%	0,4%	0,5%
LUSC	2,3%	0,8%	1,2%	0,8%	0,4%
<b>Câncer uterino</b>					
UCEC	3,3%	2,1%	2,1%	2,1%	2,4%
UCS	5,1%	0,0%	0,0%	2,5%	1,3%

**Legenda:** Carcinoma Colorretal (CRC); Adenocarcinoma de Estômago (STAD); Câncer de Mama (BRCA); Adenocarcinoma de Próstata (PRAD); Adenocarcinoma Pulmonar (LUAD); Carcinoma de pulmão de células escamosas (LUSC); Carcinoma Endometrial do Corpo Uterino (UCEC); Carcinossarcoma Uterino (UCS).



**Figura 12.** Análise de sobrevida global da família *SMYD* em relação à mutação (Kaplan-Meier plotter). **(A)** Adenocarcinoma de cólon; **(B)** Câncer de pulmão de células escamosas. Mutações (vermelho) e tipo selvagem (preto). ( $P < 0,05$  com significância estatística).

#### 4.2. Avaliação do número de cópias alteradas

A alteração do número de cópias (CNA) refere-se a alterações nos números de cópias do DNA que ocorrem em locais específicos do genoma. Essas alterações podem levar à ativação de oncogenes ou à supressão de genes supressores tumorais. Compreender a funcionalidade deste mecanismo é crucial para avançar no desenvolvimento de potenciais marcadores terapêuticos e diagnósticos (98). Os dados obtidos do CNA para os genes da família *SMYD* através do cBioPortal foram analisados usando o algoritmo GISTIC (99).

No adenocarcinoma colorretal, a alteração mais frequente foi o ganho de baixo nível, exceto *SMYD4*, que exibiu frequência de 55,5% de perda heterozigótica (Figura 12 A). No adenocarcinoma de estômago, foram observadas amplificações no *SMYD2* (2,2%) e *SMYD3* (3,7%). No entanto, a alteração mais frequente observada foi o ganho de baixo nível em todos os genes, exceto pelo *SMYD4*, que apresentou uma perda heterozigótica de 37,4%. Além disso, *SMYD3* e *SMYD4* exibiram uma perda homozigótica de 1,4% (Figura 12 B).

Em relação ao adenocarcinoma de próstata, o *SMYD2* obteve uma frequência de amplificação de 1,5%, enquanto os demais genes apresentaram frequências de amplificação inferiores a 1%. *SMYD1* e *SMYD5* apresentaram maior frequência de perda heterozigótica (7,6% e 10,1%, respectivamente) em comparação com ganho de

baixo nível. *SMYD4* exibiu o maior nível de perda heterozigótica, com 22,9%, enquanto *SMYD2* e *SMYD3* apresentaram a maior frequência de ganho de baixo nível (15,2% e 11%, respectivamente) (Figura 12 C).

Na avaliação dos subtipos de câncer de pulmão, o carcinoma de pulmão de células escamosas apresentou uma frequência de amplificação de 1,8% para *SMYD2*, 2,6% para *SMYD3* e 2,2% para *SMYD5*, enquanto os demais genes apresentaram frequências inferiores a 1%. *SMYD4* exibiu perda heterozigótica como alteração mais frequente (60,9%), e os demais genes apresentaram predominantemente ganho de baixo nível, ultrapassando 40% (Figura 12 D). No adenocarcinoma pulmonar, *SMYD2* e *SMYD3* apresentaram frequências de amplificação de 3,4% e 2,7%, respectivamente. A alteração mais comum entre os genes da família *SMYD* foi o ganho de baixo nível, exceto pelo *SMYD4*, que teve uma frequência de perda heterozigótica de 47,4% (Figura 12 E).

As frequências de amplificação foram notáveis no carcinoma de mama invasivo, com 18,6% e 19,2% para *SMYD2* e *SMYD3*, respectivamente, enquanto o ganho de baixo nível permaneceu predominante, ultrapassando 40%. *SMYD1* exibiu perda heterozigótica como alteração mais frequente (8,2%), enquanto ganho de baixo nível foi responsável por 7,9% das alterações. O *SMYD5* apresentou uma distribuição entre ganho de baixo nível e perda heterozigótica, com frequências de 8,3% e 8,4%, respectivamente. Curiosamente, mais uma vez o *SMYD4* apresentou a perda heterozigótica como a alteração mais comum, atingindo 51,3% dos casos (Figura 12 F).

No carcinoma endometrial, o *SMYD2*, *SMYD3* e *SMYD4* exibiram frequências de amplificação de 2,4%, 3,1% e 1,3%, respectivamente. Embora a alteração de maior frequência entre os genes tenha sido o ganho de baixo nível, o *SMYD4* apresentou como maior frequência a perda heterozigótica atingindo 21,0% dos casos (Figura 12 G).

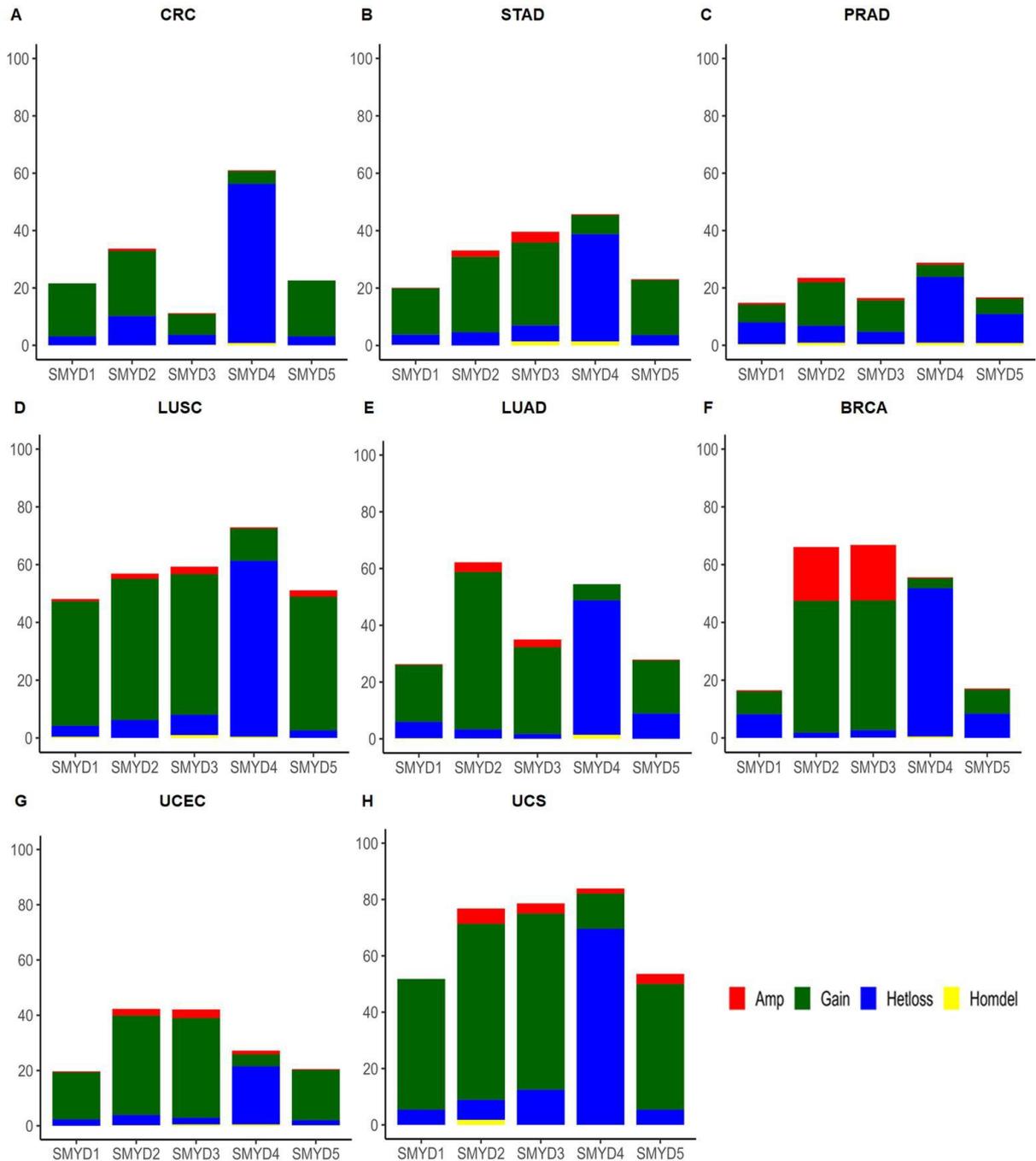
No caso do carcinosarcoma uterino, todos os genes exibiram graus variados de amplificação, com frequências de 5,4% para *SMYD2*, 3,6% para *SMYD3*, 1,8% para *SMYD4* e 3,6% para *SMYD5*. Apesar dessas amplificações, o ganho de baixo nível continuou sendo a alteração mais comum observada entre os genes, sendo responsável por mais de 40% dos casos, exceto pelo *SMYD4* que exibiu uma

frequência de 69,9% para perda heterozigótica. Além disso, um pequeno subconjunto de pacientes (1,8%) apresentou perda homozigótica para *SMYD2* (Figura 12 H).

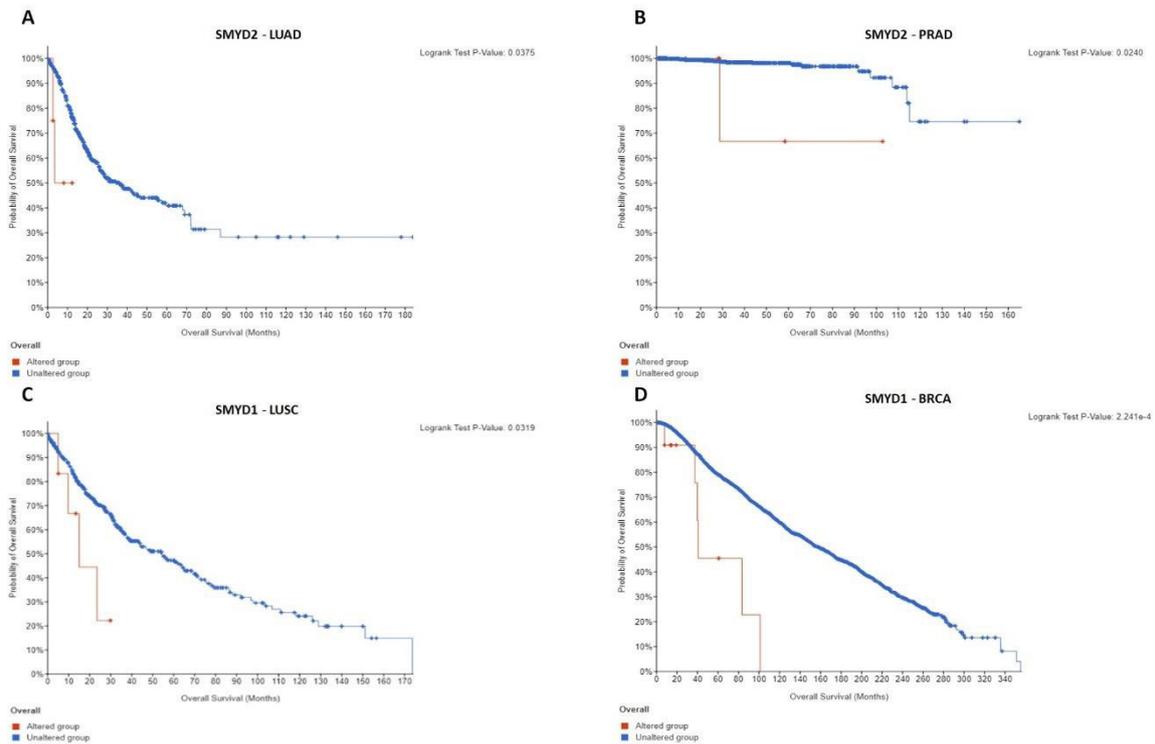
Notavelmente, o *SMYD4* demonstrou um padrão de alteração consistente em todos os tumores examinados, caracterizado por uma alta frequência de perda heterozigótica e pela ocorrência de deleção homozigótica em alguns pacientes. Esses achados são consistentes com estudos anteriores que sugerem uma ação em potencial de supressor tumoral para o *SMYD4* (82,83,100).

Em seguida, foi investigado se as alterações observadas no CNA se correlacionariam com a OS. Foi encontrado associações significativas indicando uma pior OS para pacientes com *SMYD2* em STAD ( $p = 0,0375$ ) e PRAD ( $p = 0,0240$ ). Relação semelhante foi encontrada para pacientes do CNA para o *SMYD1* em LUSC ( $p = 0,0319$ ) e em BRCA ( $p = 2,241e-4$ ) (Figura 13 A-D). Apesar do baixo número de pacientes na coorte de grupos alterados, esses resultados sugerem que há uma diferença significativa na sobrevida global, indicando um mau prognóstico para o grupo alterado em comparação ao grupo inalterado.

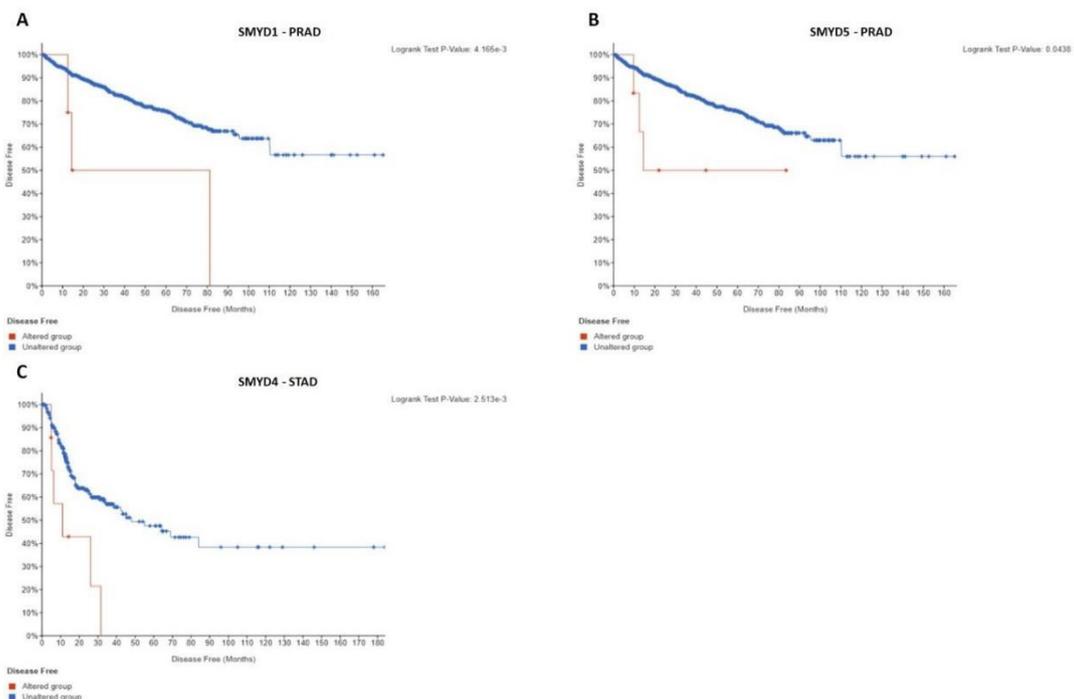
Em relação à sobrevida livre de doença (DSF), foram observadas associações significativas para CNA em *SMYD1* ( $p = 4,166e-3$ ) e *SMYD5* ( $p = 0,0438$ ) em PRAD, bem como CNA no *SMYD4* ( $p = 2,513e-3$ ) em STAD (Figura 14 A-C). Estes resultados indicam que quando o número de cópias destes genes é alterado, há uma diferença significativa na DSF que resulta em um mau prognóstico.



**Figura 13.** Frequência de alterações no número de cópias na família *SMYD*. (A) Adenocarcinoma colorretal – n = 616 para *SMYD1*, 2, 4 e 5; n = 1943 para *SMYD3*. (B) Adenocarcinoma de estômago – n = 589. (C) Adenocarcinoma de próstata – n = 2.848 para *SMYD* 1, 2 e 5; n = 3890 para *SMYD3*; n = 1835 para *SMYD4*. (D) Carcinoma de pulmão de células escamosas – n = 501. (E) Adenocarcinoma de pulmão – n = 1.109 para *SMYD1*, 2 e 4; n = 1973 para *SMYD3*; n = 1158 para *SMYD5*. (F) Câncer de mama invasivo – n = 3.469. (G) Carcinoma de endométrio – n = 620. (H) Carcinossarcoma uterino – n = 56. (Amp – amplificação de alto nível; Gain – ganho de baixo nível; Hetloss – deleção heterozigótica; Homdel – deleção homozigótica).



**Figura 14.** Sobrevida global da família *SMYD* em relação ao CNA (portal cBio). (A) Adenocarcinoma de estômago (Grupo alterado: 4 pacientes. Grupo inalterado: 527 pacientes). (B) Adenocarcinoma de próstata (Grupo alterado: 6 pacientes. Grupo inalterado: 794 pacientes). (C) Carcinoma de pulmão de células escamosas (Grupo alterado: 6 pacientes. Grupo inalterado: 489 pacientes). (D) Câncer de mama invasivo (Grupo alterado: 11 pacientes. Grupo inalterado: 3.048). Grupo alterado (vermelho) e grupo inalterado (azul) ( $P < 0,05$  com significância estatística).



**Figura 15.** Sobrevida livre de doença da família *SMYD* em relação ao CNA (portal cBio). (A) Adenocarcinoma de próstata (Grupo alterado: 4 pacientes. Grupo inalterado: 693 pacientes). (B) Adenocarcinoma de próstata (Grupo alterado: 6 pacientes. Grupo inalterado: 961 pacientes) (C) Adenocarcinoma de estômago (Grupo alterado: 7 pacientes. Grupo inalterado: 345 pacientes). Grupo alterado (vermelho) e grupo inalterado (azul) ( $P < 0,05$  com significância estatística).

#### 4.3. Perfil de expressão do mRNA

A expressão do mRNA foi realizada usando a opção em “Z-score” em relação aos dados de amostras diploides obtidos para cada tipo de câncer através do cBioPortal. As análises de mapa de calor e de diagrama de caixa (boxplot, em inglês) foram usadas para avaliar o nível de expressão da família *SMYD*. O câncer colorretal obteve uma coorte de 382 pacientes, e o *SMYD1* e *SMYD4* exibiram níveis de baixa expressão. Por outro lado, *SMYD2*, *SMYD3* e *SMYD5* apresentaram uma tendência a alta expressão (Figura 15 A).

No adenocarcinoma de estômago, a expressão de *SMYD1* permaneceu a nível basal. *SMYD2*, *SMYD3* e *SMYD5* apresentaram alta expressão na maioria dos pacientes. Mais uma vez, o *SMYD4* apresentou uma tendência de baixa expressão (Figura 15 B). Já no adenocarcinoma de próstata, o *SMYD1* manteve um nível de expressão basal, enquanto todos os outros genes da família *SMYD* foram observados com baixa expressão na maioria das amostras avaliadas (Figura 15 C).

No carcinoma de pulmão de células escamosas, o *SMYD1* e *SMYD3* apresentaram um nível de expressão basal. Já o *SMYD2* e *SMYD5* apresentaram-se com alta expressão e, mais uma vez, *SMYD4* estava com baixa expressão (Figura 15 D). Em contrapartida, o adenocarcinoma de pulmão apresentou um padrão de expressão distinto. Enquanto o *SMYD1* exibiu um nível basal, os genes *SMYD2*, *SMYD3* e *SMYD5* apresentaram níveis de alta expressão. O *SMYD4* foi encontrado, mais uma vez, com baixa expressão na maioria dos pacientes (Figura 15 E).

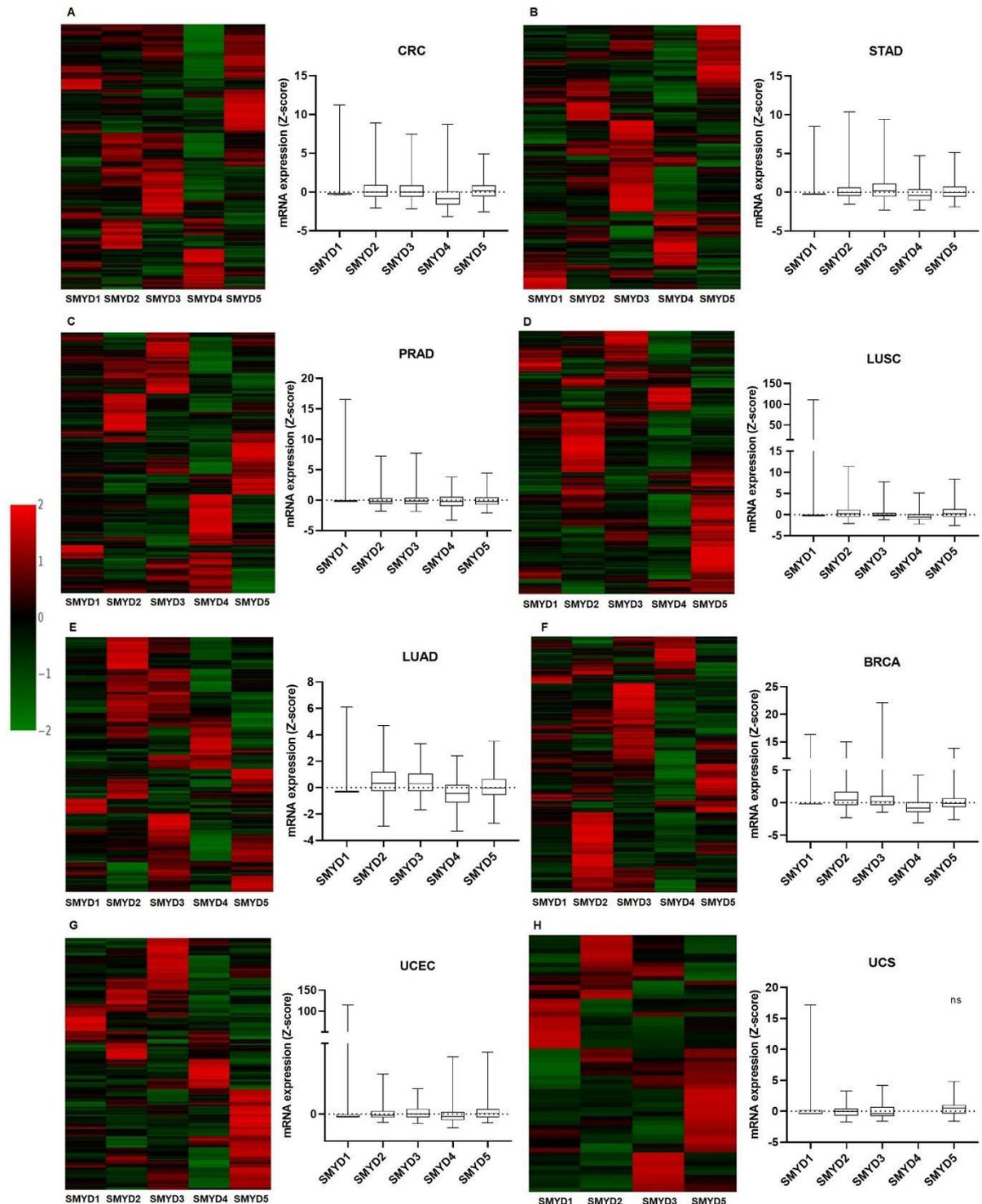
Em pacientes com câncer de mama invasivo, *SMYD1*, *SMYD4* e *SMYD5* exibiram um baixo nível de expressão. Notavelmente, tanto o *SMYD3* quanto o *SMYD2* apresentaram-se altamente expressos (Figura 15 F).

No carcinoma endometrial, *SMYD1*, *SMYD2* e *SMYD4* apresentaram-se baixo expressos, enquanto o *SMYD3* e *SMYD5* apresentaram-se alto expressos (Figura 15 G). Por fim, os dados de expressão de mRNA sobre carcinosarcoma uterino foram

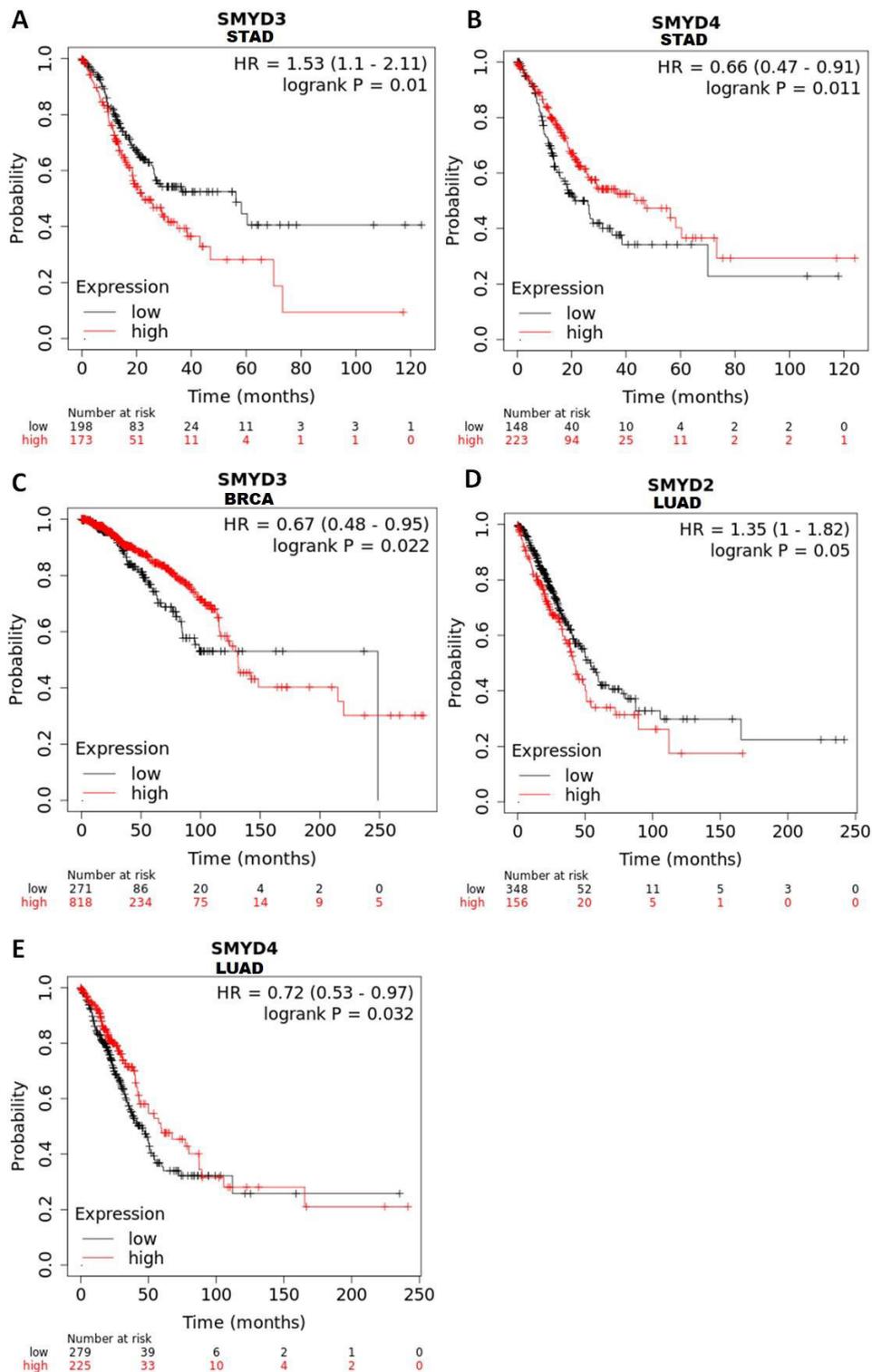
escassos. No entanto, em uma coorte com 57 pacientes, o *SMYD1*, *SMYD2* e *SMYD3* exibiram um baixo nível de expressão. Já o *SMYD5* exibiu alta expressão na maioria dos pacientes. Não havia dados disponíveis para o *SMYD4* (Figura 15 H).

Em todas as coortes de tumores avaliados, o *SMYD4* seguiu um padrão de baixa expressão. Esse resultado associa-se com os achados do CNA, em que apresentou uma alta frequência de perda heterozigótica. O *SMYD1* permaneceu com baixo nível de expressão, exceto em PRAD, STAD e LUAD, onde apresentou-se a níveis basais de expressão. A baixa expressão do *SMYD1* não está alinhada com os achados no CNA. Além disso, a maior frequência de mutação do *SMYD1* entre os genes da família *SMYD* não se reflete ao nível transcricional.

Por fim, para avaliar se há uma correlação entre a sobrevida global e o nível de expressão do mRNA na família *SMYD* como fator prognóstico, o banco de dados Kaplan-Meier foi utilizado para obter as informações disponíveis sobre STAD, BRCA e LUAD. Os grupos foram separados entre pacientes com alta e baixa expressão de cada gene em seus respectivos tumores. Foi descoberto que a alta expressão do *SMYD3* está relacionada a uma pior sobrevida global em STAD e BRCA ( $p = 0,01$  e  $p = 0,02$ , respectivamente). Além disso, a alta expressão de *SMYD2* está relacionada a uma pior sobrevida geral entre os pacientes com LUAD ( $p < 0,05$ ). Por outro lado, a baixa expressão de *SMYD4* foi associada a uma OS reduzida em pacientes com STAD e LUAD. (Figura 16 A-E).



**Figura 16.** Expressão de mRNA da família SMYD em tumores sólidos. (A) Câncer colorretal (n= 382). (B) Adenocarcinoma de estômago (n= 415). (C) Adenocarcinoma de próstata (*SMYD1*, 4 e 5: n= 498; *SMYD2*: n= 486; *SMYD3*: n= 493). (D) Câncer de pulmão de células escamosas (n= 501). (E) Adenocarcinoma de pulmão (n= 169). (F) Câncer de mama invasivo (1100). (G) Carcinoma endometrial do corpo uterino (n= 177). (H) Carcinosarcoma uterino (n= 57). A baixa expressão foi definida como um valor abaixo de 50 percentil e a alta expressão foi definida como um valor acima de 50 percentil.



**Figura 17.** Análise de sobrevida da família *SMYD* (gráfico de Kaplan-Meier). (A-B) Adenocarcinoma de estômago. (C) Câncer de mama invasivo. (D-E) Adenocarcinoma de pulmão. Alta expressão (vermelho) e baixa expressão (preto). ( $P < 0,05$  com significância estatística).

#### 4.3.1. Correlação do mRNA entre os membros da família SMYD

A análise de correlação de Pearson foi utilizada para analisar se existe uma correlação dos níveis de expressão do mRNA entre os cinco membros da família *SMYD*. No câncer colorretal, observou-se que há uma correlação positiva significativa entre o *SMYD2* e *SMYD3* ( $p = 2,357e-022$ ), bem como entre *SMYD2* e *SMYD5* ( $p = 4,055e-005$ ). Uma correlação negativa foi observada entre o *SMYD3* e *SMYD4* ( $p = 0,008$ ) e entre *SMYD4* e *SMYD5* ( $p = 3,015e-004$ ) (Figura 17 A).

No adenocarcinoma de estômago, a correlação mostrou que embora o *SMYD1* tenha permanecido a nível basal, o *SMYD2* ( $p = 9,068e-005$ ) e o *SMYD3* ( $p = 0,001$ ) estavam altamente expressos, indicando assim uma correlação negativa. Por outro lado, foi observada uma correlação positiva entre *SMYD2* e *SMYD3* ( $p = 2,087e-016$ ), *SMYD2* e *SMYD5* ( $1,994e-014$ ) e *SMYD3* e *SMYD5* ( $p = 4,143e-007$ ). (Figura 17 B).

A análise de correlação indicou que existe uma correlação positiva entre o *SMYD2* e *SMYD3* ( $p = 0,029$ ), bem como entre *SMYD3* e *SMYD5* ( $p = 7,957e-005$ ). Por outro lado, foram encontradas correlações negativas entre *SMYD2* e *SMYD4* ( $p = 0,002$ ), *SMYD3* e *SMYD4* ( $p = 1,761e-008$ ) e, finalmente, entre *SMYD4* e *SMYD5* ( $p = 1,309e-009$ ) no adenocarcinoma de próstata (Figura 17 C).

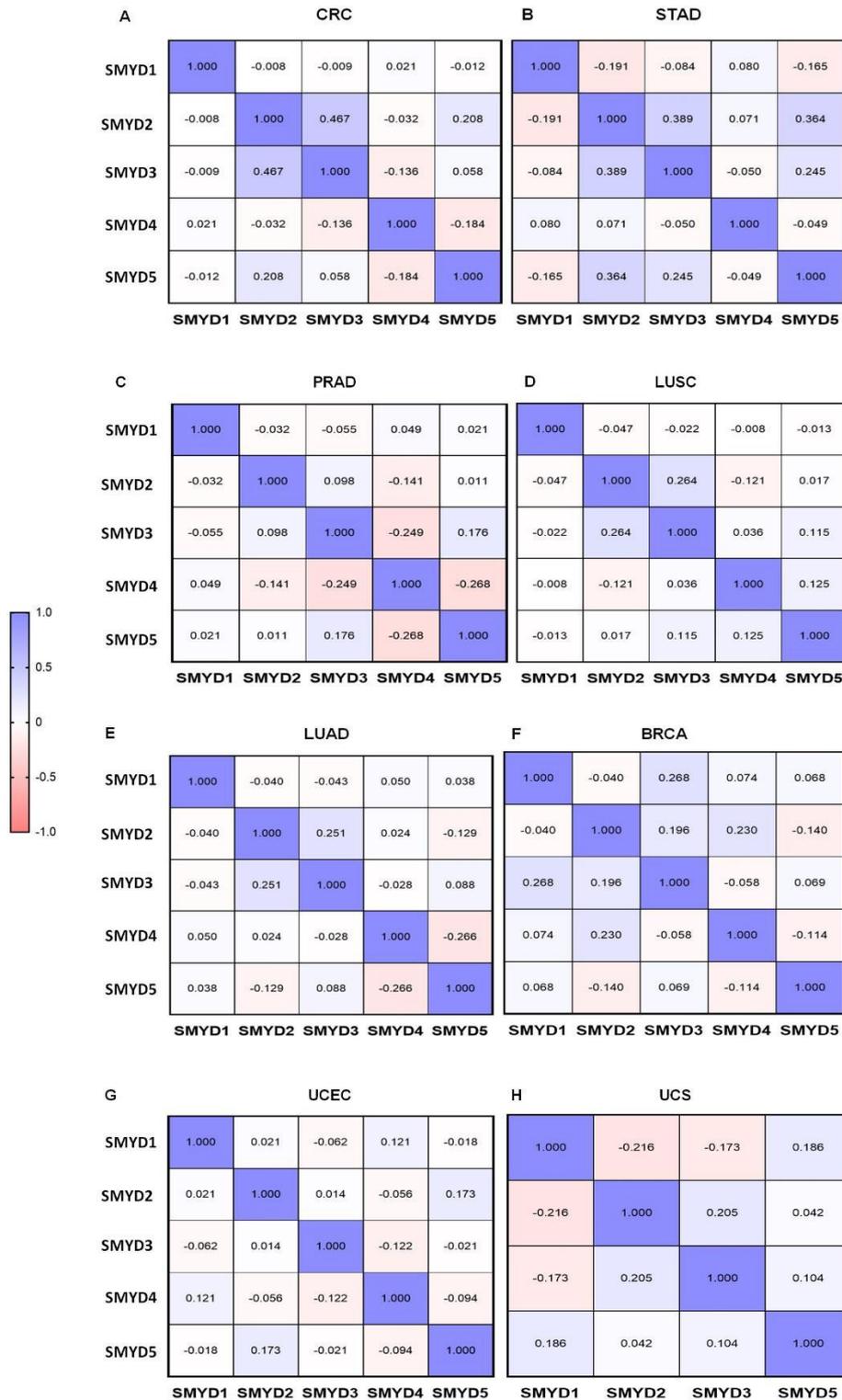
Foram encontradas correlações positivas no carcinoma de pulmão de células escamosas entre o *SMYD2* e *SMYD3* ( $p = 1,900e-009$ ), *SMYD3* e *SMYD5* ( $p = 0,010$ ), bem como entre *SMYD4* e *SMYD5* ( $p = 0,005$ ). Por fim, foi observada uma correlação negativa entre *SMYD2* e *SMYD4* ( $p = 0,007$ ) (Figura 17 D). Já no adenocarcinoma de pulmão, uma correlação positiva foi encontrada entre *SMYD2* e *SMYD3* ( $p = 0,001$ ). Observou-se uma correlação negativa entre *SMYD4* e *SMYD5* ( $p = 4,666e-004$ ). (Figura 17 E).

No câncer de mama invasivo, houve uma correlação positiva entre o *SMYD1* e *SMYD4* ( $p = 6,045e-005$ ) e entre o *SMYD2* e *SMYD5* ( $p = 8,466e-009$ ). Ocorreu uma correlação negativa entre *SMYD1* e *SMYD3* ( $p = 0,040$ ), *SMYD3* e *SMYD4* ( $p = 4,967e-005$ ) e *SMYD4* e *SMYD5* ( $0,002$ ). (Figura 17 F). Como *SMYD2* e *SMYD3* apresentaram a maior frequência de amplificação entre os tumores avaliados, foi feita uma correlação entre a coorte que continha o CNA amplificado e o mRNA com alta expressão. Ambos apresentaram uma correlação positiva significativa ( $p = 0,0001$ ) (Tabela 3).

**Tabela 4.** Correlação entre a expressão de CNA e mRNA no câncer de mama.

Gene	Pacientes (n)	Pearson	Valor de p
<i>SMYD2</i>	301	0.3398	<0.0001
<i>SMYD3</i>	281	0.2207	<0.0001

Foi observada uma correlação positiva entre *SMYD1* e *SMYD3* ( $p = 3,061e-004$ ), bem como entre *SMYD2* e *SMYD3* ( $p = 0,009$ ) e entre *SMYD2* e *SMYD4* ( $p = 0,002$ ) no carcinoma endometrial (Figura 17 G). Já para o carcinossarcoma uterino não houve correlação estatisticamente significativa na expressão de mRNA dentro da família *SMYD* na coorte analisada (Figura 17 H).



**Figura 18.** Correlação da expressão de mRNA da família SMYD em tumores sólidos. (A) Câncer colorretal (n= 382). (B) Adenocarcinoma de estômago (n= 415). (C) Adenocarcinoma de próstata (*SMYD1*, 4 e 5: n= 498; *SMYD2*: n= 486; *SMYD3*: n= 493). (D) Câncer de pulmão de células escamosas (n= 501). (E) Adenocarcinoma de pulmão (n= 169). (F) Câncer de mama invasivo (1100). (G) Carcinoma endometrial do corpo uterino (n= 177). (H) Carcinosarcoma uterino (n= 57). O teste de correlação de Pearson foi usado para comparação do nível de expressão entre os cinco genes em cada tumor avaliado ( $p = 0,05$ ).

## 5. DISCUSSÃO

Este estudo analisou as alterações moleculares da família *SMYD*, como mutação, alteração do número de cópias, expressão de mRNA e o impacto prognóstico dessas alterações no tempo de sobrevida entre os tumores sólidos mais incidentes. A análise *in silico* provou ser uma ferramenta essencial para o avanço dos estudos moleculares relacionados à carcinogênese, tendo surgido como um instrumento indispensável para explorar as vulnerabilidades relacionadas ao câncer, tornando-se um auxílio essencial para o desenvolvimento e o direcionamento de novos estudos na área. A análise das alterações genéticas mais predominantes nos genes da família *SMYD* revelou distinções importantes entre cada membro, sua relação com a progressão do câncer e sua utilidade como ferramentas de prognóstico.

No contexto das mutações, a mutação do tipo missense surge como a mais frequente observada nos genes *SMYD*. O domínio SET conservado, em particular, abriga um total de 107 mutações missense ao contabilizar em todos os cinco genes da família *SMYD*. Entretanto, nenhuma dessas mutações demonstraram uma predominância clara, o que torna difícil determinar seu possível impacto sobre a dinâmica funcional da família *SMYD*.

O *SMYD1* difere de outros *SMYDs* por apresentar uma expressão predominante nos tecidos cardíacos, assumindo um papel crucial no desenvolvimento cardíaco embrionário e sua sustentação durante a vida adulta (101). Foi descrito que o *SMYD1* forma um complexo com o skNAC, uma proteína presente nos tecidos esqueléticos e cardíacos responsável pela regeneração dos tecidos (102). Entretanto, a função do *SMYD1* durante o processo de carcinogênese precisa ser melhor elucidada. No presente estudo, o *SMYD1* foi o gene mais mutado e a expressão do mRNA permaneceu a nível basal nos tumores STAD, PRAD e LUAD, e mostrou uma tendência à *downregulação* nos tumores CRC, LUSC, BRCA, UCEC e UCS. Esses achados sugerem não haver relação com o CNA, uma vez que o ganho de baixo nível foi a alteração mais frequente, com exceção dos tumores PRAD e BRCA, nos quais o *SMYD1* apresentou uma perda heterozigótica em maior frequência.

Os achados para o *SMYD1* em PRAD implicam que as alterações do CNA poderiam servir como fator prognóstico para a sobrevida livre de doença (DFS). Zhang et al. (2019) também observaram uma relação semelhante, demonstrando que a

amplificação do *SMYD1* estava ligada a uma diminuição da sobrevida livre de progressão (PFS) (62). Embora o CNA comumente identificado no *SMYD1* tenha sido de perda heterozigótica, seu potencial como biomarcador merece atenção, necessitando de mais estudos para avaliar de forma abrangente o papel do *SMYD1* na carcinogênese da próstata.

Por outro lado, embora a alteração mais frequente no CNA do *SMYD1* no BRCA correspondesse à perda heterozigótica, um subconjunto de onze pacientes com amplificação se correlacionou com uma pior sobrevida global em comparação com o grupo sem alterações. Embora o estudo de Jianping et al. (2019) tenha sido relacionado à expressão de mRNA e não ao CNA, eles mostraram que o *SMYD1* estava com baixo nível de expressão no câncer de mama invasivo e que o aumento da expressão se correlacionou com uma boa sobrevida livre da doença (DFS) (59). Apesar da aparente contradição de ambos os resultados, isso pode ser possível devido ao fato de que a alteração do CNA nem sempre leva a alterações no mRNA (103). Notavelmente, o *SMYD1* pode influenciar no BRCA e mais estudos são necessários para determinar se o *SMYD1* pode servir como um biomarcador prognóstico para o BRCA.

Na família *SMYD*, o *SMYD2* é um dos membros mais estudados associado ao câncer. Ele apresentou uma baixa frequência de mutação, sendo o menos mutado entre os cinco membros (0,4%). Entretanto, com relação ao CNA, todos os tumores avaliados apresentaram ganho de baixo nível como a alteração mais frequente. Particularmente no BRCA, o *SMYD2* foi amplificado em 18,6% dos casos. Nos tumores CRC, STAD, LUAD, LUSC e BRCA, a expressão do mRNA também foi elevada, em concordância com os achados do CNA. No entanto, o *SMYD2* mostrou-se baixo expresso na maioria dos pacientes com PRAD, UCEC e UCS, que não foi de acordo com os achados do CNA.

Na análise de sobrevida, a alta expressão do *SMYD2* no LUAD foi associada à redução do tempo de sobrevida, correspondendo aos achados na análise de mRNA. Estudos relacionaram a superexpressão do *SMYD2* ao desenvolvimento do tumor, estando associada à proliferação, migração e invasão celular. A superexpressão associou-se também à metástase no câncer colorretal (104,105), adenocarcinoma de estômago (61,106) e adenocarcinoma de pulmão (68), e foi correlacionada à progressão do câncer de mama e também à invasão e metástase (66,107). Embora o

*SMYD2* tenha sido associado à progressão de tumores ginecológicos (108), não há descrição sobre seu envolvimento no carcinoma endometrial do corpo uterino e no carcinossarcoma uterino.

No PRAD, o perfil de CNA do *SMYD2* exibiu uma prevalência notável de ganho de baixo nível, mas na coorte de análise de sobrevida, o subgrupo com alteração de deleção homozigótica exibiu uma sobrevida global (OS) inferior. Em relação à sobrevida livre de progressão (PFS), Zhang et al. (2019) mostraram que a amplificação do *SMYD2* estava associada à sua redução (62). Portanto, embora essas duas situações sejam diferentes, isso sugere que o *SMYD2* tem uma influência na progressão da doença e na sobrevida do paciente. Elucidar seu mecanismo durante esse processo é importante para direcionar e aprimorar a eficácia do tratamento em pacientes com PRAD.

No STAD, a alteração do *SMYD2* no CNA também foi associada a um pior OS. Entretanto, até o momento, há apenas um estudo que avaliou a expressão do mRNA do *SMYD2* no tumor em questão. Embora tenha sido associado a um aumento da expressão de *SMYD2* no STAD, não houve associação com a OS, mas houve uma correlação com outros tumores do sistema digestivo (61). No entanto, o *SMYD2* em geral, possui ação oncogênica e sua superexpressão está associada ao desenvolvimento tumoral e pode servir como um potencial biomarcador terapêutico.

Em relação ao *SMYD3*, que também é amplamente estudado e associado ao câncer, surgiu como o segundo gene mais mutado neste estudo, com uma taxa de mutação de 0,6%. Em termos de alterações do número de cópias, todos os tumores avaliados apresentaram predominantemente ganhos de baixo nível. No entanto, a amplificação do *SMYD3* no câncer de mama foi notável, representando 19,2% das alterações. Os níveis de expressão de mRNA variaram entre os diferentes tipos de tumores: a alta expressão foi observada em CRC, STAD, LUAD, BRCA e UCEC, enquanto o *SMYD3* em PRAD e UCS mostrou uma tendência à baixa expressão. O carcinoma de pulmão de células escamosas, por outro lado, manteve-se a um nível basal de expressão.

A análise de sobrevida mostrou que a alta expressão do *SMYD3* em STAD e BRCA estava correlacionada com a redução do tempo de sobrevida. A análise de sobrevivência relacionada à mutação mostrou que houve uma redução no tempo de

sobrevida em pacientes com câncer de cólon quando o *SMYD3* sofreu mutação. No entanto, o banco de dados Kaplan-meier plotter usado para essa análise não especifica qual tipo de mutação estava presente na coorte selecionada. Até o momento, também não há estudos que correlacionem mutações no *SMYD3* com a OS no câncer de cólon.

A superexpressão do *SMYD3* já foi relacionada a vários tipos de tumores, contribuindo com o desenvolvimento e a progressão do câncer, regulando a proliferação, apoptose, invasão celular e os processos metastáticos (71). Estudos anteriores associaram a superexpressão de *SMYD3* ao crescimento do CRC e a um prognóstico desfavorável (72,109), à migração celular, proliferação e à progressão do ciclo celular, além de correlacionar sua alta expressão a um pior prognóstico e tempo de sobrevida no STAD (61,110).

No LUAD, descobriu-se que o *SMYD3* estava superexpresso e associado à ativação da família oncogênica RAS, participando da proliferação celular e do crescimento do tumor (76,111). A superexpressão de *SMYD3* também foi associada à proliferação celular no câncer de mama podendo ser vantajoso para o crescimento do tumor, além de poder servir como um possível biomarcador terapêutico (112,113).

Embora não tenha sido observado a tendência à alta expressão do *SMYD3* na coorte do PRAD, pesquisas anteriores associaram sua alta expressão ao aumento da proliferação celular e à progressão do câncer de próstata. Foi relatado que o *SMYD3* está envolvido na ativação dos receptores de andrógenos e pode servir como fator prognóstico (62,114,115). Portanto, são necessários mais estudos para entender melhor a função do *SMYD3* no processo cancerígeno do PRAD.

No presente estudo, o *SMYD4* foi o terceiro gene mais mutado, sendo responsável por 0,5% do total de mutações. Em todos os tumores avaliados, observou-se uma alta frequência de perda heterozigótica, o que se alinha com os achados do mRNA em todos os tipos de tumores analisados, os quais exibiram uma tendência à baixa expressão. Essa tendência pode ocorrer por conta da localização do *SMYD4*, que se encontra na região cromossômica 17p13.3, descrita na literatura por sofrer perda heterozigótica em vários tipos de câncer (82,116–119).

A análise de sobrevida revelou que a baixa expressão do *SMYD4* em STAD e LUAD associou-se à redução do tempo de sobrevida. Curiosamente, um estudo

conduzido por Liu et al. (2020) constatou que o *SMYD4* estava superexpresso na maioria dos pacientes com STAD, um achado que contradiz nossos resultados. Uma possível explicação para esse fato se dá pela capacidade do *SMYD4* se ligar e elevar a expressão de Nanog em células tronco cancerígenas (84), auxiliando a proliferação, migração e invasão celular. Portanto, é evidente que o *SMYD4* desempenha um papel no STAD, ressaltando a necessidade de mais pesquisas para entender melhor seu mecanismo de ação nesse tipo de tumor (120,121).

Os achados no câncer de mama foram consistentes com estudos anteriores, do qual associavam a baixa expressão do *SMYD4* ao desenvolvimento do tumor mamário (82,83). Em relação ao adenocarcinoma de próstata, um estudo conduzido pelo Zhang et. al. (2019) mostrou que o *SMYD4* estava baixo expresso quando comparado a tecidos não cancerígenos, o que entrou em acordo com o achado nesse estudo (62). A função do *SMYD4* no processo de carcinogênese ainda é pouco elucidada, porém, há grandes indícios que o mesmo possua função de supressor tumoral (82,83,100), mas pode agir também como oncogene em células tronco cancerígenas ao se ligar a receptores Nanog quando hiperexpresso, contribuindo para o desenvolvimento tumoral (84).

O *SMYD5* foi o segundo gene mais mutado, juntamente com o *SMYD3* (0,6%). No CNA, o gene teve uma frequência maior de ganho de baixo nível nos tumores CRC, STAD, LUAD, LUSC, UCEC e UCS, enquanto o PRAD teve a perda heterozigótica mais frequente. O BRCA ficou dividido entre ganho de baixo nível (8,3%) e perda heterozigótica (8,4%). Em termos de mRNA, todos os tumores apresentaram-se com alta expressão exceto pelo PRAD e BRCA, que tenderam à baixa expressão.

Um estudo realizado por Kidder et al. (2017) mostrou que, no câncer de cólon, no carcinoma de pulmão de células escamosas e câncer de mama, a perda do *SMYD5* levou à superexpressão de genes envolvidos em sua carcinogênese, o que designa ao *SMYD5* uma possível função de supressor tumoral (90). No entanto, esses dados apenas corroboram os achados do câncer de mama no presente estudo.

No STAD, essa análise corroborou estudos anteriores em que a alta expressão do *SMYD5* associou-se ao desenvolvimento do adenocarcinoma de estômago (61,122). No entanto, os achados sobre a expressão do mRNA no PRAD contradizem

os achados de outro estudo, ressaltando a necessidade de pesquisas adicionais para esclarecer o papel do *SMYD5* nesse tipo de câncer (62). Quanto aos outros tumores, nenhum estudo existente estabelece uma relação com o *SMYD5*. Isso destaca a importância de uma melhor compreensão dos mecanismos de ação do *SMYD5* no processo carcinogênico.

Na análise de correlação de Pearson, alguns achados foram dignos de nota devido a alterações no padrão de expressão da família *SMYD* em alguns tumores. No adenocarcinoma de próstata, apesar da baixa expressão do *SMYD2-5*, observou-se uma associação negativa entre *SMYD2* e *SMYD4*, *SMYD3* e *SMYD4*, e *SMYD4* e *SMYD5*. Isso foi possível porque um número significativo de pacientes apresentou padrões de expressão opostos. Por exemplo, alguns pacientes demonstraram altos níveis de expressão no *SMYD2* (19,8%), *SMYD3* (19,5%), *SMYD4* (27,1%) e *SMYD5* (19,3%). Apesar da tendência à baixa expressão, esse grupo de pacientes poderiam se beneficiar de inibidores para os genes mencionados acima (56).

Embora o *SMYD2* e *SMYD5* estejam predominantemente alto expressos, o *SMYD4* esteja com baixa expressão e o *SMYD3* mantenha um nível basal no carcinoma de células escamosas do pulmão, foi observado uma correlação positiva entre eles. Essa correlação é possível porque alguns pacientes apresentam o mesmo padrão de expressão entre esses genes, apesar da correlação ser relativamente fraca.

Por outro lado, houve uma correlação significativamente positiva entre o *SMYD2* e *SMYD5*, que foram encontrados com alta e baixa expressão no BRCA, respectivamente, apesar de ser uma correlação fraca. Isso pode ocorrer porque alguns pacientes apresentaram padrões de expressão semelhantes. Houve uma correlação negativa entre *SMYD4* e *SMYD5*, embora ambos tenham sido encontrados baixo expressos. Isso pode ter ocorrido porque alguns pacientes apresentaram padrões de expressão opostos. Vale a pena observar que o *SMYD4* e *SMYD5* apresentaram 10,5% e 14,9%, respectivamente, de aumento na expressão do mRNA, o que demonstra a importância da estratificação entre os pacientes com câncer de mama para melhor direcionamento terapêutico com base nas características biológicas e moleculares de cada paciente (123).

Curiosamente, *SMYD2* e *SMYD3*, que tiveram uma alta frequência de amplificação de CNA, foram positivamente correlacionados com a alta expressão do

mRNA. Isso sugere que o aumento do CNA pode levar ao aumento da expressão de mRNA, indicando que alterações dessas características podem servir como possíveis biomarcadores para o câncer de mama.

Na UCEC, o *SMYD1* e *SMYD2* apresentaram-se baixo expressos, enquanto o *SMYD3* apresentou-se com aumento de sua expressão. Mesmo assim, houve uma correlação positiva entre *SMYD1* e *SMYD3*, e entre *SMYD2* e *SMYD3*. Isso possivelmente ocorreu devido ao fato de que o *SMYD3* esteve com baixa expressão em 9,6% dos pacientes.

Portanto, as análises de correlação ressaltam um ponto importante: o papel crítico da medicina personalizada no tratamento contemporâneo do câncer (43). Isso é destacado pela observação de que, mesmo que um gene exiba predominantemente um determinado padrão de expressão, ele pode se manifestar de maneira oposta em um subconjunto de pacientes com o mesmo tumor. Isso resalta a complexidade da expressão gênica e a necessidade de abordagens terapêuticas individualizadas.

## 6. CONCLUSÃO

Ao avaliar as alterações moleculares da família *SMYD* nos tumores mais incidentes na população brasileira em busca de potenciais biomarcadores terapêuticos e diagnósticos, pode-se concluir:

- No *SMYD1*, alterações do CNA no PRAD e BRCA associaram-se a redução no tempo de sobrevida livre da doença e na sobrevida global, respectivamente.
- Alta expressão do *SMYD2* no LUAD levou a redução no tempo de sobrevida global, enquanto que alterações no CNA associaram-se à redução do tempo de sobrevida global em PRAD e STAD. O *SMYD2* tem ação de oncogene e está envolvido em processos que influenciam o desenvolvimento tumoral.
- Alta expressão do *SMYD3* teve correlação com a diminuição do tempo de sobrevida no STAD e BRCA. O *SMYD3* possui também ação de oncogene auxiliando o desenvolvimento tumoral.
- O *SMYD4* se destacou entre os demais genes por seguir um mesmo padrão de alteração no CNA, apresentando uma alta frequência de perda heterozigótica, e baixa expressão do mRNA em todos os tumores avaliados nesse

estudo. Sua baixa expressão associou-se à redução da sobrevida no STAD e LUAD, indicando uma possível ação de supressor tumoral.

- O *SMYD5* obteve alterações no CNA associando-se a redução do tempo de sobrevida livre da doença no PRAD, sendo indicativo de mau prognóstico.
- A família *SMYD*, portanto, pode servir como potenciais biomarcadores prognósticos, sendo que o *SMYD2* e *SMYD3* podem atuar também como biomarcadores terapêuticos nos tumores citados acima.
- As alterações nos padrões de expressão entre a família *SMYD* indicados pela análise de correlação de Pearson nos tumores aqui avaliados, ressaltou a importância do desenvolvimento e implementação da medicina personalizada durante o tratamento oncológico, ao direcionar as abordagens terapêuticas de acordo com a necessidade de cada paciente.

## 7. REFERÊNCIAS

1. What Is Cancer? - NCI [Internet]. [cited 2023 Apr 18]. Available from: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>
2. Robert A. Weinberg. The Biology of Cancer. Second edition. Garland Science, Taylor & Francis Group;
3. Hallmarks of Cancer: The Next Generation - ScienceDirect [Internet]. [cited 2023 Jun 27]. Available from: <https://www.sciencedirect.com/science/article/pii/S0092867411001279>
4. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* 2022 Jan 12;12(1):31–46.
5. Cell Signaling by Receptor Tyrosine Kinases. *Cell.* 2010 Jun 25;141(7):1117–34.
6. Stromal fibroblasts in cancer initiation and progression - PubMed [Internet]. [cited 2023 Jun 28]. Available from: <https://pubmed.ncbi.nlm.nih.gov/15549095/>
7. Cree IA. Cancer biology. *Methods Mol Biol Clifton NJ.* 2011;731:1–11.
8. The Bcl-2 apoptotic switch in cancer development and therapy | Oncogene [Internet]. [cited 2023 Jun 30]. Available from: <https://www.nature.com/articles/1210220>
9. Fulda S. Tumor resistance to apoptosis. *Int J Cancer.* 2009;124(3):511–5.
10. Blasco MA. Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet.* 2005 Aug;6(8):611–22.
11. Ferrara N. Pathways mediating VEGF-independent tumor angiogenesis. *Cytokine Growth Factor Rev.* 2010 Feb 1;21(1):21–6.
12. Cavallaro U, Christofori G. Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nat Rev Cancer.* 2004 Feb;4(2):118–32.
13. Loose D, Van de Wiele C. The Immune System and Cancer. *Cancer Biother Radiopharm.* 2009 Jun;24(3):369–76.
14. Annibaldi A, Widmann C. Glucose metabolism in cancer cells. *Curr Opin Clin Nutr Metab Care.* 2010 Jul;13(4):466–70.
15. Lh W, Cf W, N R, Yk S. Loss of Tumor Suppressor Gene Function in Human Cancer: An Overview. *Cell Physiol Biochem Int J Exp Cell Physiol Biochem Pharmacol* [Internet]. 2018 [cited 2023 Jul 9];51(6). Available from: <https://pubmed.ncbi.nlm.nih.gov/30562755/>
16. THOMPSON & THOMPSON. Genética médica. 7ª Edição.

17. Role of Oncogenes and Tumor-suppressor Genes in Carcinogenesis: A Review - PubMed [Internet]. [cited 2023 Jul 9]. Available from: <https://pubmed.ncbi.nlm.nih.gov/33109539/>
18. The MYC oncogene - the grand orchestrator of cancer growth and immune evasion - PubMed [Internet]. [cited 2023 Jul 9]. Available from: <https://pubmed.ncbi.nlm.nih.gov/34508258/>
19. Sun W, Yang J. Functional Mechanisms for Human Tumor Suppressors. *J Cancer*. 2010;1:136.
20. M L, R R, G A, Lk L, L LC. Metabolic functions of the tumor suppressor p53: Implications in normal physiology, metabolic disorders, and cancer. *Mol Metab* [Internet]. 2020 Mar [cited 2023 Jul 10];33. Available from: <https://pubmed.ncbi.nlm.nih.gov/31685430/>
21. Cancer Epigenetics: An Overview - ScienceDirect [Internet]. [cited 2023 Aug 25]. Available from: <https://www.sciencedirect.com/science/article/pii/S0188440922001424>
22. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis*. 2010 Jan;31(1):27–36.
23. Zhang Y, Sun Z, Jia J, Du T, Zhang N, Tang Y, et al. Overview of Histone Modification. *Adv Exp Med Biol*. 2021;1283:1–16.
24. INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Estimativa 2023 : incidência de câncer no Brasil. Rio de Janeiro : INCA, 2022;
25. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. *Nat Rev Dis Primer*. 2019 Sep 23;5(1):66.
26. Sarhangi N, Hajjari S, Heydari SF, Ganjizadeh M, Rouhollah F, Hasanzad M. Breast cancer in the era of precision medicine. *Mol Biol Rep*. 2022 Oct 1;49(10):10023–37.
27. Provenzano E, Ulaner GA, Chin SF. Molecular Classification of Breast Cancer. *PET Clin*. 2018 Jul 1;13(3):325–38.
28. Wang G, Zhao D, Spring DJ, DePinho RA. Genetics and biology of prostate cancer. *Genes Dev*. 2018 Jan 9;32(17–18):1105–40.
29. Abate-Shen C, Shen MM. Molecular genetics of prostate cancer. *Genes Dev*. 2000 Jan 10;14(19):2410–34.
30. Rawla P. Epidemiology of Prostate Cancer. *World J Oncol*. 2019 Apr;10(2):63–89.
31. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol*. 2021 Oct;14(10):101174.
32. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet Lond Engl*. 2019 Oct 19;394(10207):1467–80.
33. Baidoun F, Elshiwiy K, Elkeraie Y, Merjaneh Z, Khoudari G, Sarmini MT, et al. Colorectal Cancer Epidemiology: Recent Trends and Impact on Outcomes. *Curr Drug Targets*. 2021;22(9):998–1009.

34. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe*. 2013 Aug 14;14(2):207–15.
35. Kwong TNY, Wang X, Nakatsu G, Chow TC, Tipoe T, Dai RZW, et al. Association Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. *Gastroenterology*. 2018 Aug;155(2):383-390.e8.
36. Wistuba II, Gazdar AF. Lung cancer preneoplasia. *Annu Rev Pathol*. 2006;1:331–48.
37. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer*. 2014 Aug;14(8):535–46.
38. Bade BC, Dela Cruz CS. Lung Cancer 2020: Epidemiology, Etiology, and Prevention. *Clin Chest Med*. 2020 Mar;41(1):1–24.
39. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. *Lancet Lond Engl*. 2016 Nov 26;388(10060):2654–64.
40. Lauren P. THE TWO HISTOLOGICAL MAIN TYPES OF GASTRIC CARCINOMA: DIFFUSE AND SO-CALLED INTESTINAL-TYPE CARCINOMA. AN ATTEMPT AT A HISTO-CLINICAL CLASSIFICATION. *Acta Pathol Microbiol Scand*. 1965;64:31–49.
41. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet Lond Engl*. 2020 Aug 29;396(10251):635–48.
42. Johnson CA, James D, Marzan A, Armaos M. Cervical Cancer: An Overview of Pathophysiology and Management. *Semin Oncol Nurs*. 2019 Apr 1;35(2):166–74.
43. Bell DW, Ellenson LH. Molecular Genetics of Endometrial Carcinoma. *Annu Rev Pathol*. 2019 Jan 24;14:339–67.
44. Boggess JF, Kilgore JE, Tran AQ. 85 - Uterine Cancer. In: Niederhuber JE, Armitage JO, Kastan MB, Doroshow JH, Tepper JE, editors. *Abeloff's Clinical Oncology (Sixth Edition)* [Internet]. Philadelphia: Elsevier; 2020 [cited 2023 Apr 18]. p. 1508-1524.e4. Available from: <https://www.sciencedirect.com/science/article/pii/B9780323476744000852>
45. Felix AS, Brinton LA. Cancer Progress and Priorities: Uterine Cancer. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol*. 2018 Sep;27(9):985–94.
46. Albert M, Helin K. Histone methyltransferases in cancer. *Semin Cell Dev Biol*. 2010 Apr;21(2):209–20.
47. Black JC, Van Rechem C, Whetstine JR. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell*. 2012 Nov 30;48(4):491–507.
48. Husmann D, Gozani O. Histone lysine methyltransferases in biology and disease. *Nat Struct Mol Biol*. 2019 Oct;26(10):880–9.

49. Yang C, Zhang J, Ma Y, Wu C, Cui W, Wang L. Histone methyltransferase and drug resistance in cancers. *J Exp Clin Cancer Res CR*. 2020 Aug 28;39(1):173.
50. Alvarez-Venegas R, Avramova Z. SET-domain proteins of the Su(var)3-9, E(z) and trithorax families. *Gene*. 2002 Feb 20;285(1–2):25–37.
51. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001 Aug 10;293(5532):1074–80.
52. Chetverina DA, Lomaev DV, Erokhin MM. Polycomb and Trithorax Group Proteins: The Long Road from Mutations in *Drosophila* to Use in Medicine. *Acta Naturae*. 2020;12(4):66–85.
53. Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*. 2010 Sep 17;142(6):967–80.
54. Leinhart K, Brown M. SET/MYND Lysine Methyltransferases Regulate Gene Transcription and Protein Activity. *Genes*. 2011 Feb 21;2(1):210–8.
55. Spellmon N, Holcomb J, Trescott L, Sirinupong N, Yang Z. Structure and Function of SET and MYND Domain-Containing Proteins. *Int J Mol Sci*. 2015 Jan 8;16(1):1406–28.
56. Padilla A, Manganaro JF, Huesgen L, Roess DA, Brown MA, Crans DC. Targeting Epigenetic Changes Mediated by Members of the SMYD Family of Lysine Methyltransferases. *Molecules*. 2023 Feb 20;28(4):2000.
57. Gottlieb PD, Pierce SA, Sims RJ, Yamagishi H, Weihe EK, Harriss JV, et al. Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. *Nat Genet*. 2002 May;31(1):25–32.
58. Coda DM, Lingua MF, Morena D, Foglizzo V, Bersani F, Ala U, et al. SMYD1 and G6PD modulation are critical events for miR-206-mediated differentiation of rhabdomyosarcoma. *Cell Cycle Georget Tex*. 2015;14(9):1389–402.
59. Song J, Liu Y, Chen Q, Yang J, Jiang Z, Zhang H, et al. Expression patterns and the prognostic value of the SMYD family members in human breast carcinoma using integrative bioinformatics analysis. *Oncol Lett*. 2019 Apr;17(4):3851–61.
60. Peveling-Oberhag J, Wolters F, Döring C, Walter D, Sellmann L, Scholtysik R, et al. Whole exome sequencing of microdissected splenic marginal zone lymphoma: a study to discover novel tumor-specific mutations. *BMC Cancer*. 2015 Oct 24;15:773.
61. Liu D, Wang X, Shi E, Wang L, Nie M, Li L, et al. Comprehensive Analysis of the Value of SMYD Family Members in the Prognosis and Immune Infiltration of Malignant Digestive System Tumors. *Front Genet*. 2021;12:699910.
62. Zhang Y, Yan L, Yao W, Chen K, Xu H, Ye Z. Integrated Analysis of Genetic Abnormalities of the Histone Lysine Methyltransferases in Prostate Cancer. *Med Sci Monit Int Med J Exp Clin Res*. 2019 Jan 7;25:193–239.

63. Brown MA, Sims RJ, Gottlieb PD, Tucker PW. Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Mol Cancer*. 2006 Jun 28;5:26.
64. The Tale of Two Domains - Molecular & Cellular Proteomics [Internet]. [cited 2023 Dec 24]. Available from: [https://www.mcponline.org/article/S1535-9476\(20\)31215-9/fulltext](https://www.mcponline.org/article/S1535-9476(20)31215-9/fulltext)
65. Huang J, Perez-Burgos L, Placek BJ, Sengupta R, Richter M, Dorsey JA, et al. Repression of p53 activity by Smyd2-mediated methylation. *Nature*. 2006 Nov 30;444(7119):629–32.
66. Nakakido M, Deng Z, Suzuki T, Dohmae N, Nakamura Y, Hamamoto R. Dysregulation of AKT Pathway by SMYD2-Mediated Lysine Methylation on PTEN. *Neoplasia N Y N*. 2015 Apr;17(4):367–73.
67. Reynoird N, Mazur PK, Stellfeld T, Flores NM, Lofgren SM, Carlson SM, et al. Coordination of stress signals by the lysine methyltransferase SMYD2 promotes pancreatic cancer. *Genes Dev*. 2016 Apr 1;30(7):772–85.
68. Wu L, Kou F, Ji Z, Li B, Zhang B, Guo Y, et al. SMYD2 promotes tumorigenesis and metastasis of lung adenocarcinoma through RPS7. *Cell Death Dis*. 2021 May 2;12(5):439.
69. Li LX, Zhou JX, Calvet JP, Godwin AK, Jensen RA, Li X. Lysine methyltransferase SMYD2 promotes triple negative breast cancer progression. *Cell Death Dis*. 2018 Feb 27;9(3):326.
70. Sun JJ, Li HL, Ma H, Shi Y, Yin LR, Guo SJ. SMYD2 promotes cervical cancer growth by stimulating cell proliferation. *Cell Biosci*. 2019;9:75.
71. Rueda-Robles A, Audano M, Álvarez-Mercado AI, Rubio-Tomás T. Functions of SMYD proteins in biological processes: What do we know? An updated review. *Arch Biochem Biophys*. 2021 Nov 15;712:109040.
72. Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, et al. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol*. 2004 Aug;6(8):731–40.
73. Foreman KW, Brown M, Park F, Emtage S, Harriss J, Das C, et al. Structural and functional profiling of the human histone methyltransferase SMYD3. *PLoS One*. 2011;6(7):e22290.
74. Van Aller GS, Reynoird N, Barbash O, Huddleston M, Liu S, Zmoos AF, et al. Smyd3 regulates cancer cell phenotypes and catalyzes histone H4 lysine 5 methylation. *Epigenetics*. 2012 Apr;7(4):340–3.
75. Kunizaki M, Hamamoto R, Silva FP, Yamaguchi K, Nagayasu T, Shibuya M, et al. The lysine 831 of vascular endothelial growth factor receptor 1 is a novel target of methylation by SMYD3. *Cancer Res*. 2007 Nov 15;67(22):10759–65.
76. Mazur PK, Reynoird N, Khatri P, Jansen PWTC, Wilkinson AW, Liu S, et al. SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature*. 2014 Jun 12;510(7504):283–7.

77. Li J, Zhao L, Pan Y, Ma X, Liu L, Wang W, et al. SMYD3 overexpression indicates poor prognosis and promotes cell proliferation, migration and invasion in non-small cell lung cancer. *Int J Oncol*. 2020 Sep;57(3):756–66.
78. Lin F, Wu D, Fang D, Chen Y, Zhou H, Ou C. STAT3-induced SMYD3 transcription enhances chronic lymphocytic leukemia cell growth in vitro and in vivo. *Inflamm Res Off J Eur Histamine Res Soc Al*. 2019 Sep;68(9):739–49.
79. Zhu CL, Huang Q. Overexpression of the SMYD3 Promotes Proliferation, Migration, and Invasion of Pancreatic Cancer. *Dig Dis Sci*. 2020 Feb;65(2):489–99.
80. Xiao D, Wang H, Hao L, Guo X, Ma X, Qian Y, et al. The roles of SMYD4 in epigenetic regulation of cardiac development in zebrafish. *PLoS Genet*. 2018 Aug;14(8):e1007578.
81. Chen K, Chen Z, Wu D, Zhang L, Lin X, Su J, et al. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat Genet*. 2015 Oct;47(10):1149–57.
82. Hu L, Zhu YT, Qi C, Zhu YJ. Identification of Smyd4 as a potential tumor suppressor gene involved in breast cancer development. *Cancer Res*. 2009 May 1;69(9):4067–72.
83. Han S, Zou H, Lee JW, Han J, Kim HC, Cheol JJ, et al. miR-1307-3p Stimulates Breast Cancer Development and Progression by Targeting SMYD4. *J Cancer*. 2019;10(2):441–8.
84. Liu S, Cheng K, Zhang H, Kong R, Wang S, Mao C, et al. Methylation Status of the Nanog Promoter Determines the Switch between Cancer Cells and Cancer Stem Cells. *Adv Sci Weinh Baden-Wurttemberg*. 2020 Mar;7(5):1903035.
85. Nørøxe DS, Yde CW, Østrup O, Michaelsen SR, Schmidt AY, Kinalis S, et al. Genomic profiling of newly diagnosed glioblastoma patients and its potential for clinical utility – a prospective, translational study. *Mol Oncol*. 2020 Nov;14(11):2727–43.
86. Kidder BL, Hu G, Cui K, Zhao K. SMYD5 regulates H4K20me3-marked heterochromatin to safeguard ES cell self-renewal and prevent spurious differentiation. *Epigenetics Chromatin*. 2017;10:8.
87. Aljazi MB, Gao Y, Wu Y, He J. SMYD5 is a histone H3-specific methyltransferase mediating mono-methylation of histone H3 lysine 36 and 37. *Biochem Biophys Res Commun*. 2022 Apr 9;599:142–7.
88. Stender JD, Pascual G, Liu W, Kaikkonen MU, Do K, Spann NJ, et al. Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. *Mol Cell*. 2012 Oct 12;48(1):28–38.
89. Fujii T, Tsunesumi SI, Sagara H, Munakata M, Hisaki Y, Sekiya T, et al. Smyd5 plays pivotal roles in both primitive and definitive hematopoiesis during zebrafish embryogenesis. *Sci Rep*. 2016 Jul 5;6:29157.
90. Kidder BL, He R, Wangsa D, Padilla-Nash HM, Bernardo MM, Sheng S, et al. SMYD5 Controls Heterochromatin and Chromosome Integrity during Embryonic Stem Cell Differentiation. *Cancer Res*. 2017 Dec 1;77(23):6729–45.

91. Meng X, Zhao Y, Liu J, Wang L, Dong Z, Zhang T, et al. Comprehensive analysis of histone modification-associated genes on differential gene expression and prognosis in gastric cancer. *Exp Ther Med*. 2019 Sep;18(3):2219–30.
92. Gao H, Chakraborty G, Lee-Lim AP, Mavrakis KJ, Wendel HG, Giancotti FG. Forward genetic screens in mice uncover mediators and suppressors of metastatic reactivation. *Proc Natl Acad Sci U S A*. 2014 Nov 18;111(46):16532–7.
93. SMYD5 acts as a potential biomarker for hepatocellular carcinoma - PubMed [Internet]. [cited 2023 Dec 28]. Available from: <https://pubmed.ncbi.nlm.nih.gov/35218722/>
94. SAKAMOTO, Luis Henrique Toshihiro. Papel de metiltransferases de proteínas no desenvolvimento e prognóstico de leucemias linfóides agudas da infância. Universidade de Brasília; 2014.
95. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*. 2019 Jan 8;47(D1):D941–7.
96. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012 May;2(5):401–4.
97. Chandrashekar DS, Karthikeyan SK, Korla PK, Patel H, Shovon AR, Athar M, et al. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia N Y N*. 2022 Mar;25:18–27.
98. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. *Nat Genet*. 2003 Aug;34(4):369–76.
99. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12(4):R41.
100. Zhang Y, Wang H, Li C, Gao L, Zheng Y, Chang W, et al. CircSMYD4 regulates proliferation, migration and apoptosis of hepatocellular carcinoma cells by sponging miR-584-5p. *Cancer Cell Int*. 2020 Nov 19;20(1):556.
101. Tracy C, Warren JS, Szulik M, Wang L, Garcia J, Makaju A, et al. The Smyd Family of Methyltransferases: Role in Cardiac and Skeletal Muscle Physiology and Pathology. *Curr Opin Physiol*. 2018 Feb;1:140–52.
102. Berkholz J, Orgeur M, Stricker S, Munz B. skNAC and Smyd1 in transcriptional control. *Exp Cell Res*. 2015 Aug 15;336(2):182–91.
103. Bhattacharya A, Bense RD, Urzúa-Traslaviña CG, de Vries EGE, van Vugt MATM, Fehrmann RSN. Transcriptional effects of copy number alterations in a large set of human cancers. *Nat Commun*. 2020 Feb 5;11(1):715.
104. Meng F, Liu X, Lin C, Xu L, Liu J, Zhang P, et al. SMYD2 suppresses APC2 expression to activate the Wnt/ $\beta$ -catenin pathway and promotes epithelial-mesenchymal transition in colorectal cancer. *Am J Cancer Res*. 2020;10(3):997–1011.

105. Pan L, Fan Y, Zhou L. SMYD2 epigenetically activates MEX3A and suppresses CDX2 in colorectal cancer cells to augment cancer growth. *Clin Exp Pharmacol Physiol*. 2022 Sep;49(9):959–69.
106. Komatsu S, Ichikawa D, Hirajima S, Nagata H, Nishimura Y, Kawaguchi T, et al. Overexpression of SMYD2 contributes to malignant outcome in gastric cancer. *Br J Cancer*. 2015 Jan 20;112(2):357–64.
107. Zeng Y, Qiu R, Yang Y, Gao T, Zheng Y, Huang W, et al. Regulation of EZH2 by SMYD2-Mediated Lysine Methylation Is Implicated in Tumorigenesis. *Cell Rep*. 2019 Nov 5;29(6):1482-1498.e4.
108. Zhang X, Wang Y. Identification of hub genes and key pathways associated with the progression of gynecological cancer. *Oncol Lett*. 2019 Dec;18(6):6516–24.
109. Liu N, Sun S, Yang X. Prognostic significance of stromal SMYD3 expression in colorectal cancer of TNM stage I-III. *Int J Clin Exp Pathol*. 2017;10(8):8901–7.
110. Wang L, Wang QT, Liu YP, Dong QQ, Hu HJ, Miao Z, et al. ATM Signaling Pathway Is Implicated in the SMYD3-mediated Proliferation and Migration of Gastric Cancer Cells. *J Gastric Cancer*. 2017 Dec;17(4):295–305.
111. Alnabulsi SM, Al-Shar'i NA. Hit identification of SMYD3 enzyme inhibitors using structure-based pharmacophore modeling. *Future Med Chem*. 2019 May;11(10):1107–17.
112. Hamamoto R, Silva FP, Tsuge M, Nishidate T, Katagiri T, Nakamura Y, et al. Enhanced SMYD3 expression is essential for the growth of breast cancer cells. *Cancer Sci*. 2006 Feb;97(2):113–8.
113. Fenizia C, Bottino C, Corbetta S, Fittipaldi R, Floris P, Gaudenzi G, et al. SMYD3 promotes the epithelial-mesenchymal transition in breast cancer. *Nucleic Acids Res*. 2019 Feb 20;47(3):1278–93.
114. Vieira FQ, Costa-Pinheiro P, Ramalho-Carvalho J, Pereira A, Menezes FD, Antunes L, et al. Deregulated expression of selected histone methylases and demethylases in prostate carcinoma. *Endocr Relat Cancer*. 2014 Feb;21(1):51–61.
115. Liu C, Wang C, Wang K, Liu L, Shen Q, Yan K, et al. SMYD3 as an oncogenic driver in prostate cancer by stimulation of androgen receptor transcription. *J Natl Cancer Inst*. 2013 Nov 20;105(22):1719–28.
116. Jenal M, Britschgi C, Fey MF, Tschan MP. Inactivation of the hypermethylated in cancer 1 tumour suppressor--not just a question of promoter hypermethylation? *Swiss Med Wkly*. 2010;140:w13106.
117. Wu GJ, Shan XN, Li MF, Shi SL, Zheng QP, Yu L, et al. Preliminary study on the loss of heterozygosity at 17p13 in gastric and colorectal cancers. *World J Gastroenterol*. 1997 Sep 15;3(3):160–2.
118. Tsuchiya E, Tanigami A, Ishikawa Y, Nishida K, Hayashi M, Tokuchi Y, et al. Three New Regions on Chromosome 17p13.3 Distal to p53 with Possible Tumor Suppressor Gene Involvement in Lung Cancer. *Jpn J Cancer Res Gann*. 2000 Jun;91(6):589–96.

119. Park SY, Kang YS, Kim BG, Lee SH, Lee ED, Lee KH, et al. Loss of heterozygosity on the short arm of chromosome 17 in uterine cervical carcinomas. *Cancer Genet Cytogenet.* 1995 Jan;79(1):74–8.
120. Lin T, Ding YQ, Li JM. Overexpression of Nanog protein is associated with poor prognosis in gastric adenocarcinoma. *Med Oncol.* 2012 Jun 1;29(2):878–85.
121. Ma X, Wang B, Wang X, Luo Y, Fan W. NANOGP8 is the key regulator of stemness, EMT, Wnt pathway, chemoresistance, and other malignant phenotypes in gastric cancer cells. *PLOS ONE.* 2018 Apr 24;13(4):e0192436.
122. Reyes DA, Sarría VMS, Salazar-Viedma M, D'Afonseca V. Histone Methyltransferases Useful in Gastric Cancer Research. *Cancer Inform.* 2021;20:11769351211039862.
123. Goutsouliak K, Veeraraghavan J, Sethunath V, De Angelis C, Osborne CK, Rimawi MF, et al. Towards personalized treatment for early stage HER2-positive breast cancer. *Nat Rev Clin Oncol.* 2020 Apr;17(4):233–50.

## CAPÍTULO 2

### RESUMO

As leucemias são um conjunto de doenças de característica maligna que afetam células sanguíneas, levando a alta taxa proliferativa de glóbulos brancos não funcionais impedindo o desenvolvimento hematopoiético normal. São caracterizadas pelo tipo celular afetado e pelo seu grau de maturação, classificando-se em quatro subtipos principais: Leucemia mieloide crônica (LMC), leucemia mieloide aguda (LMA), leucemia linfóide crônica (LLC) e leucemia linfóide aguda (LLA), sendo essa última o objeto de interesse do presente estudo. A LLA é caracterizada, de modo geral, pela alta proliferação de blastos, com capacidade de invadir outras regiões do corpo. Afeta em sua grande maioria crianças de 1 a 4 anos de idade.

A família *SMYD* de metiltransferases possuem até o momento poucos estudos que os correlacionem à LLA, porém já existe na literatura estudos que mostram o *SMYD2* alterado em amostras de crianças com LLA, sendo um possível indicativo prognóstico. Já o *SMYD5* foi encontrado com uma alta expressão nessas mesmas amostras e, através desse achado, o *SMYD5* foi então escolhido como objeto de estudos do presente trabalho. Foi proposto a avaliação do número de cópias alteradas (CNA) e dos níveis de expressão do mRNA da família *SMYD* de coortes obtidas através do *cBioPortal* e *BloodSpot*.

Em relação ao CNA, houve diferenças de alterações entre a família *SMYD*, sendo que o *SMYD1* e *SMYD5* apresentaram a perda heterozigótica como a alteração mais frequente, enquanto que os demais apresentaram um ganho de baixo nível como a alteração mais frequente. Já em relação a expressão do mRNA, o *SMYD1* e *SMYD2* mostram-se com baixa expressão enquanto o *SMYD5* mostrou-se superexpresso em LLA quando comparados a amostras de medula óssea saudável. Os demais membros não apresentam diferença de expressão estatisticamente significativas.

Portanto, apesar da pouca informação, a família *SMYD* tem sido correlacionada não só a LLA como também a outras malignidades sanguíneas. Estudos anteriores mostraram que o *SMYD5* possui função na hematopoiese e que estava alterado em LLA. Já foi relatado também que o *SMYD5* forma um complexo repressor com NCoR, sendo responsável pela H4K20me3 mantendo, dessa forma, a repressão de genes pró-inflamatórios. Na literatura, foi descrito que sua depleção pode levar a aberrações cromossômicas.

Dessa forma, há evidências que o *SMYD5* tenha participação na leucemogênese por estar envolvido nas funções citadas acima e, portanto, foi proposto avaliar com mais precisão os efeitos causados pelo *SMYD5* ao iniciar modelos estáveis de linhagens celulares de LLA-B expressando Cas9, para realização de knockout gênico por meio da técnica de CRISPR-Cas9.

**Palavras-chave:** Leucemias; Leucemia linfóide aguda; Família *SMYD*; *SMYD5*; CRISPR-Cas9.

## ABSTRACT

Leukemias are a group of malignant diseases that affect blood cells, leading to a high proliferation rate of non-functional white blood cells and inhibiting normal hematopoietic development. They are characterized by the type of cell affected and its stage of maturation and are classified into four main subtypes: chronic myeloid leukaemia (CML), acute myeloid leukaemia (AML), chronic lymphocytic leukaemia (CLL) and acute lymphocytic leukaemia (ALL), the last of which is the subject of this study. ALL is generally characterized by the high proliferation of blasts, with the capacity to invade other regions of the body. It mostly affects children between the ages of 1 and 4.

The SMYD family of methyltransferases has so far had few studies correlating them with ALL, but there are studies in the literature showing *SMYD2* altered in samples from children with ALL, which is a possible prognostic indicator. *SMYD5*, on the other hand, was found to be upregulated in these same samples and, as a result of this finding, *SMYD5* was chosen as the subject of this study. It was proposed to evaluate the number of altered copies (CNA) and mRNA expression levels of the SMYD family from cohorts obtained through cBioPortal and BloodSpot.

In relation to CNA, there were differences in alterations between the *SMYD* family, with *SMYD1* and *SMYD5* showing heterozygous loss as the most frequent alteration, while the others showed a low-level gain as the most frequent alteration in ALL. In terms of mRNA expression, *SMYD1* and *SMYD2* were downregulated while *SMYD5* was upregulated in ALL when compared to healthy bone marrow samples. The other members showed no statistically significant difference in expression.

Therefore, despite the limited information, the *SMYD* family has been correlated not only with ALL but also with other blood malignancies. Previous studies have shown that *SMYD5* has a function in hematopoiesis and that it was altered in ALL. It has also been reported that *SMYD5* forms a repressor complex with NCoR and is responsible for H4K20me3, thus maintaining the repression of pro-inflammatory genes. In the literature, it has been described that its depletion can lead to chromosomal aberrations.

Therefore, there is evidence that *SMYD5* plays a role in leukemogenesis due to its involvement in the above-mentioned functions. It was then proposed to evaluate the effects caused by *SMYD5* more precisely by starting stable models of ALL-B cell lines expressing Cas9, in order to perform a gene knockout using the CRISPR-Cas9 technique.

**keywords:** Leukemias; Acute lymphocytic leukemia; SMYD family; SMYD5; CRISPR-Cas9.

## 1. INTRODUÇÃO

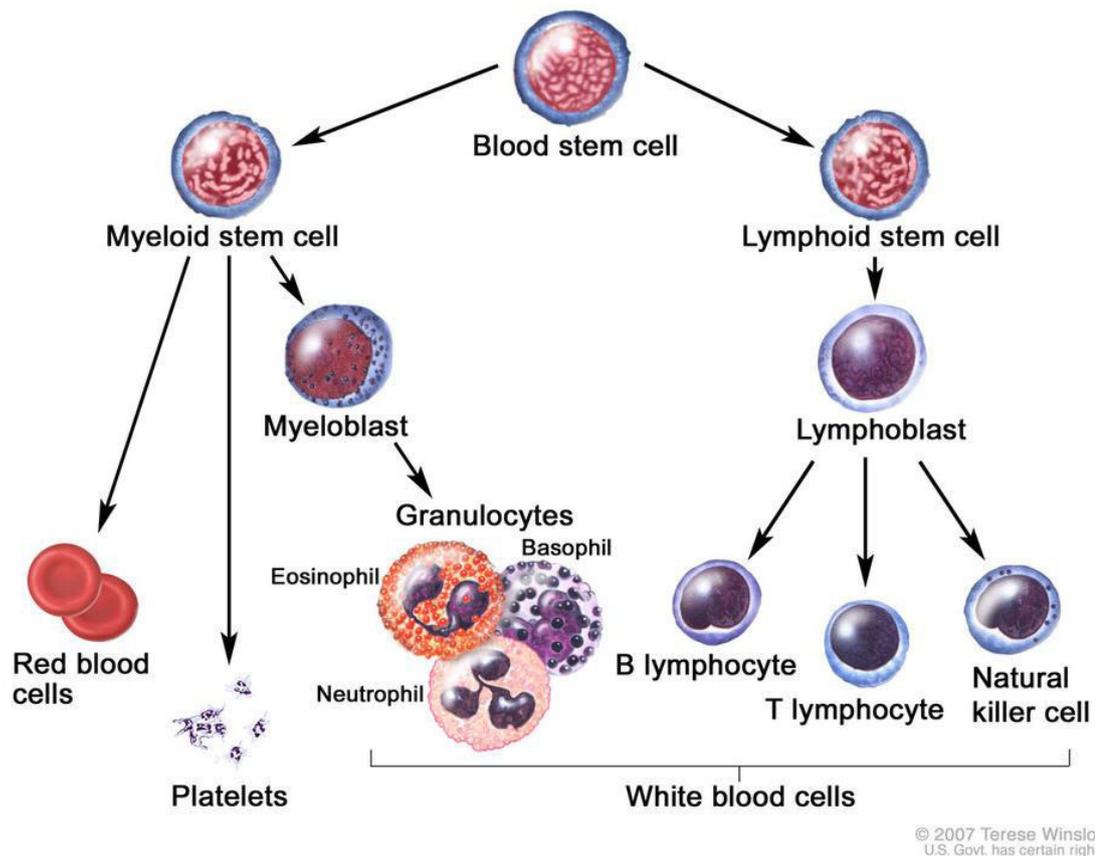
### 1.1. O que é leucemia?

As leucemias são descritas como um grupo de doenças malignas hematológicas de origem clonal de células-tronco hematopoiéticas (CTH), caracterizadas pela rápida proliferação de glóbulos brancos não funcionais, que ocupam espaço na medula óssea e impedem a produção de glóbulos vermelhos, glóbulos brancos e plaquetas normais (1,2). Essas células leucêmicas podem chegar até o sangue periférico e infiltrar outros órgãos, como por exemplo o fígado e os linfonodos (2). O processo de carcinogênese da CTH leva a formação da célula-tronco leucêmica (CTL), que apresenta danos a sua capacidade de diferenciação, além de levar ao aumento proliferativo e auto-renovação ilimitada, características essas que auxiliam na progressão da doença, recidiva e resistência a drogas terapêuticas (1).

A hematopoese é controlada através dos componentes da matriz extracelular (MEC), que formam o estroma medular e auxiliam em processos de manutenção e liberação das células hematopoiéticas para o sangue periférico. Observou-se então que, quando há alguma alteração nesse microambiente, a proliferação descontrolada de células neoplásicas ocorre através dessa interação com a MEC. A proliferação celular em células neoplásicas é dependente também da produção autócrina de citocinas que, por sua vez, auxiliam na síntese e no aumento de proteoglicanos na superfície celular, fazendo com que essas células aumentem sua capacidade de adesão à MEC (2). O microambiente exerce grande influência em processos como a auto-renovação, quiescência, diferenciação e a proliferação celular e, portanto, quando desregulado, auxilia na promoção da oncogênese. Além disso, as células leucêmicas são capazes de alterar o microambiente do qual estão inseridas, tornando-o mais permissível para a progressão da leucemia (1).

A classificação das leucemias se dá de acordo com o tipo celular que é acometido pela doença e pelo grau de maturação dessa célula. Referente ao tipo celular, podem ser classificadas como leucemias linfóides, que levam a transformação neoplásica de células progenitoras dos linfócitos do tipo T e B, e leucemias mielóides, que levam a transformação de células progenitoras responsáveis pela maturação da linhagem mieloide, como monócitos, mastócitos, basófilos, neutrófilos, eosinófilos, eritrócitos e trombócitos (Figura 1). De acordo com o grau de maturação, a

classificação se dá em leucemias agudas, que levam ao bloqueio da maturação das células sanguíneas acompanhada do aumento proliferativo dessas células imaturas, e leucemias crônicas, que são caracterizadas pela transformação parcial de células já maduras seguida pela proliferação clonal celular (1,2).



**Figura 1.** Representação do processo hematopoiético. Figura criada pelo “National Cancer Institute”. Terese Winslow (2007).

De acordo com a classificação descrita acima, as leucemias se dividem em quatro tipos principais: leucemia mieloide crônica (LMC), leucemia mieloide aguda (LMA), leucemia linfóide crônica (LLC) e leucemia linfóide aguda (LLA), sendo esta última o objeto de estudo do presente trabalho.

A LMC é caracterizada pela transformação e crescimento anormal de células mielóides em graus diferentes de maturação, sendo um subtipo raro em crianças e, por sua natureza crônica, as células afetadas geralmente funcionam adequadamente, o que permite um desenvolvimento mais brando e lento da doença quando comparado aos tipos agudos (1,3,4). A translocação nos cromossomos 9 e 22, conhecido por cromossomo filadélfia (Ph), leva a fusão gênica dos genes BCR e ABL (Bcr-Abl), e

essa alteração associa-se a LMC. Essa fusão gênica é responsável pela ativação das vias de sinalização MAPK, JAK/STAT, MYC, RAS e PI3K, que estão envolvidas em processos proliferativos, de sinalização, sobrevivência e mecanismos de controle celular (1).

A LMC quando não tratada pode progredir para fase blástica, caracterizada pelo aumento de blastos que podem ser de linhagem mieloide, linfoide ou mista/indiferenciada. Os sintomas relacionados à LMC incluem febre, fadiga e perda de peso, e com a progressão da doença, os sintomas incluem dores ósseas e sangramento. Porém, os pacientes que se encontram na fase crônica da doença são assintomáticos (3).

A LMA, por sua vez, caracteriza-se pelo descontrole no crescimento e acúmulo de mieloblastos imaturos na medula óssea que levam ao bloqueio da diferenciação celular (1,4). É uma doença agressiva que leva ao desenvolvimento de sintomas que se relacionam com falhas da medula óssea, como febre, infecção, sintomas relacionados a anemia, hematomas ou sangramentos, além da capacidade de infiltração em outros órgãos. A LMA é rara, porém atinge o segundo lugar em termo de mortalidade entre as leucemias e afeta majoritariamente pessoas da terceira idade, com a média de diagnóstico sendo de 68 anos (5).

Para fins prognósticos e melhor orientação para escolha do tratamento, recomenda-se uma análise cromossômica, e o principal propósito terapêutico é a aquisição da remissão da doença, que leva ao desaparecimento dos blastos na medula óssea e, conseqüentemente, a normalidade da produção de glóbulos brancos, glóbulos vermelhos e plaquetas (4).

A LLC resulta da proliferação exagerada e acúmulo de uma célula única de linfócito B madura CD5<sup>+</sup>, que está presente no sangue, na medula óssea, linfonodos e baço. Atinge pessoas na fase adulta, com média de diagnóstico de 70 anos de idade, porém há uma variação na incidência de acordo com a etnia e sexo biológico (6,7). A consequência clínica resultante do crescimento clonal varia, já que alguns pacientes vão a óbito em um período entre dois e três anos, e outros sobrevivem durante décadas após o diagnóstico. Essa variedade pode ocorrer tanto por fatores intrínsecos quanto por fatores extrínsecos do qual cada paciente é exposto (7).

Os fatores de risco relacionados a LLC ainda não foram identificados, entretanto, já foi observado uma pré-disposição genética que atinge cerca de 10% dos pacientes diagnosticados (6). Sabe-se também que há uma interação entre a LLC e o HIF-1 $\alpha$ , que confere plasticidade a essas células leucêmicas, além de regular o metabolismo e o ciclo celular (1).

Já a LLA será melhor descrita mais adiante. De modo geral, a LLA é resultado da proliferação descontrolada de célula linfoide B ou T, que prejudica a produção normal de glóbulos vermelhos, brancos e plaquetas (4).

Portanto, é fundamental a distinção fenotípica das células leucêmicas para fins diagnósticos, prognósticos e terapêuticos.

## 1.2. Leucemia Linfoide Aguda de células B

A leucemia linfoide aguda é caracterizada pela alta taxa de proliferação de linfócitos em estágio inicial de diferenciação (blastos), podendo invadir outros locais do corpo, inclusive regiões extramedular, que prejudica a formação normal de células sanguíneas (4,8).

A leucemia linfoide aguda de células B (LLA-B) afeta em sua grande maioria crianças de 1 a 4 anos de idade, havendo uma redução brusca ao decorrer da infância em indivíduos com idade entre 5 e 14 anos, adolescente e jovens adultos, de 15 a 39 anos de idade, com menor incidência em indivíduos entre 25 e 45 anos de idade. Nesse panorama, a sobrevida geral de pacientes acometidos pela LLA-B no período de 5 anos chega a 90% em crianças, mas em adultos com mais de 50 anos de idade há uma queda para apenas 25%. Os fatores prognósticos, portanto, englobam idade, contagem de células sanguíneas, envolvimento do sistema nervoso central, etnia, gênero e tipo celular acometido (8).

Em relação a fatores de predisposição, mesmo com o avanço no entendimento da patologia acerca da LLA, a maioria dos casos acontece em indivíduos saudáveis. Porém, existe uma associação a interações com fatores ambientais como, por exemplo, exposição a pesticidas, radiação ionizante e infecções no período gestacional ou durante a infância. Há também associação a fatores hereditários, que incluem síndromes genéticas que podem aumentar a susceptibilidade ao

desenvolvimento da LLA, como a Síndrome de Down, de Fancone e Li-Fraumeni, bem como outras anormalidades genéticas e cromossômicas que participam e afetam processos como a proliferação e diferenciação de células precursoras de linfócitos (8,9).

De modo geral, a LLA caracteriza-se por vários subtipos genéticos, com diversas alterações moleculares que incluem alterações cromossômicas como aneuploidia e rearranjos, alteração no número de cópia do DNA e mutação. Essas características impactam diretamente o prognóstico e a resposta terapêutica (9).

A aneuploidia divide-se em hiperdiploidia e hipodiploidia. A hiperdiploidia é mais comum em crianças e possui dois subtipos, sendo a alta hiperdiploidia caracterizada pela presença de 51 a 65 cromossomos, e a baixa hiperdiploidia que contém de 47 a 50 cromossomos. A alta hiperdiploidia em conjunto com algumas características clínicas, como a idade média de 4 anos e uma baixa contagem no nível de células brancas no momento do diagnóstico, associa-se a um bom prognóstico com uma taxa de 90% de sobrevivência, pelo fato de responderem bem a quimioterapia. A hipodiploidia (abaixo de 44 cromossomos), é uma condição mais rara que afeta ~7% dos pacientes com LLA, e possui um aumento em sua incidência com a idade. Possui como característica uma baixa contagem de células brancas no momento do diagnóstico e associa-se a um mau prognóstico (9).

Já os rearranjos cromossômicos levam a formação de genes quiméricos através de translocações. Entre as translocações, o rearranjo do KMT2A (conhecido também por MLL), aparece em 80% dos diagnósticos de crianças com idade de 0 a 1 ano, associando-se a um mau prognóstico (8). Outra alteração cromossômica que é comumente encontrada é a fusão dos genes BCR-ABL [t (9;22) (q34;q11)], conhecido como cromossomo filadélfia (Ph<sup>+</sup>). O Ph<sup>+</sup> aumenta com a idade sendo mais incidente em adultos de 35 a 50 anos (20% a 30% dos casos), que leva a disfunção de várias vias de sinalização, como a auto-renovação, proliferação celular e crescimento tumoral, que ocasiona um mau prognóstico (10).

Outro rearranjo comum em pacientes adolescentes e jovens adultos ocorre no DUX4, que vem associado a um bom prognóstico mesmo quando existe outro fator genético de risco. Rearranjo no MYC ocorre em cerca de 4% de pacientes adultos

com LLA-B sendo bastante raro em pacientes pediátricos, e essa característica associa-se a um mau prognóstico (11).

O rearranjo ETV6::RUNX1, que é a alteração presente na linhagem celular utilizada no presente estudo, é originada através da translocação [t (12;21) (p13;q22)], também conhecido por TEL::AML1. Afeta cerca de 25% das crianças diagnosticadas com LLA-B, sendo essa alteração a mais comum entre esses pacientes, e associa-se a resultados favoráveis com o tempo de sobrevida de 5 anos chegando a 94%, e 88% em 10 anos. Já foi identificado a presença de ETV6::RUNX1 no sangue do cordão umbilical, indicando que essa doença esteja presente durante a fase pré-natal. Essa alteração por conta própria não leva ao desenvolvimento da leucemia, sendo então necessário a presença de uma alteração genética secundária em genes associados a maturação de células B, que levará ao processo da leucemogênese (9).

#### 1.2.1. *SMYDs* e a LLA-B

A família *SMYD* é composta por cinco genes (*SMYD1-5*) dos quais possuem estudos que os associam a diversos tumores (12–18). Porém, estudos que correlacionam a família *SMYD* a malignidades em células sanguíneas e principalmente a LLA-B, ainda são escassos.

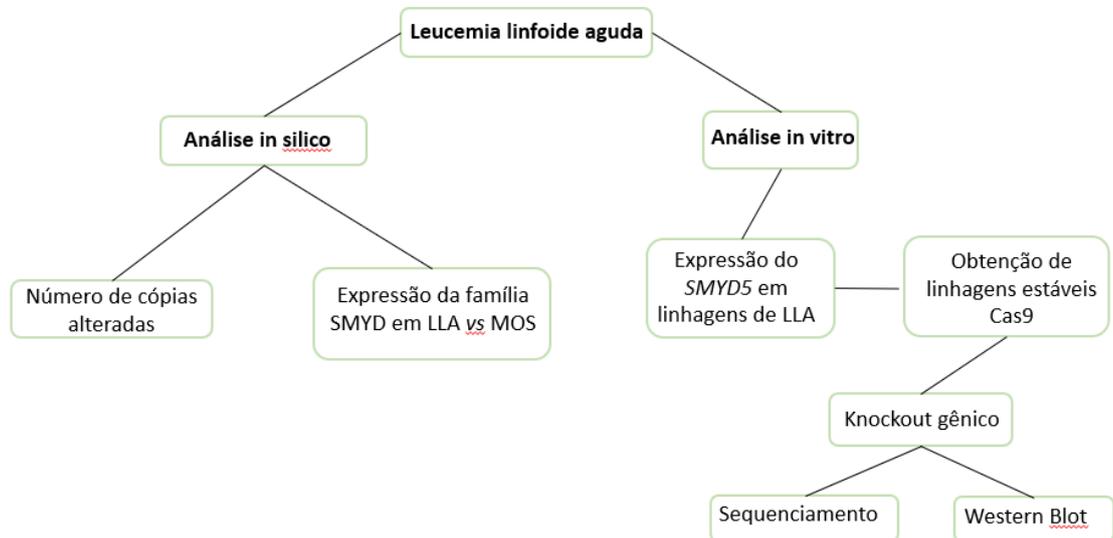
Contudo, apesar de não haver até o momento evidências da participação do *SMYD1* na LLA, mutações no *SMYD1* já foram encontradas em pacientes diagnosticados com linfoma não-Hodgkin de célula B (19).

Em relação ao *SMYD2*, foi observado que havia uma alteração de sua expressão em crianças com LLA que correlacionava-se a um mau prognóstico, indicando que o *SMYD2* poderia servir como um marcador de acompanhamento e prognóstico da doença (20). Esses resultados foram posteriormente corroborados por Zang et.al (2020) (21). Já o *SMYD3* não possui até o momento estudos que o correlacione com a LLA, mas sua alta expressão já foi descrita na leucemia linfóide crônica (22,23). *SMYD4* também não possui estudos que o relacionem a LLA.

O *SMYD5*, gene de interesse do presente estudo, apresentou alta expressão em pacientes com LLA, além de possuir funções na hematopoiese (24,25). Apesar da escassez de estudos que envolvem a participação do *SMYD5* em LLA, há informações sobre sua participação em outros tipos de câncer (13,18,26).

## 2. MATERIAIS E MÉTODOS

### 2.1. Delineamento experimental



**Figura2:** Delineamento experimental referente ao capítulo 2.

### 2.2. Base de dados online

Para análise de expressão do mRNA, a plataforma online *BloodSpot* (<https://www.fobinf.com/>) foi utilizada por conter dados sobre células hematopoiéticas saudáveis e doentes. O estudo “*Leukemia MILE*” (GSE13159) foi selecionado para comparar a expressão do *SMYD5* em amostras de leucemia linfóide aguda de células B e amostras de medula óssea saudável (27). Para análise de significância entre a expressão do gene alvo nas amostras selecionadas, o teste de Mann Whitney com valor de significância estatística de  $p < 0.05$  foi utilizado.

Para análise do número de cópia alteradas (CNA) e correlação entre mRNA e CNA, dados obtidos pelo *cBioportal* foram utilizados como descrito no capítulo 1. Para análise de correlação, os testes de Pearson e Spearman com valor de significância estatística de  $p < 0.05$  foram utilizados.

O software GraphPad prism 8.0. foi utilizado para realizar as análises estatísticas.

### 2.3. Cultivo celular

As linhagens celulares de LLA-B REH, RS4;11, Nalm-6 e 697 foram cultivadas em meio RPMI (Roswell Park Memorial Institute) 1640 (Gibco), suplementado com

bicarbonato de sódio (Sigma), penicilina e estreptomicina a 1% (Gibco Pen-Strep) e soro fetal bovino (SFB) a 10% (Gibco). As células foram mantidas em meio RPMI completo à 37°C, em estufa úmida com 5% CO<sub>2</sub>.

#### 2.4. Extração de DNA

Para extração de DNA genômico foi utilizado o protocolo alternativo de extração por Proteinase K (Biolabs #P81025), realizado da seguinte forma:

Ressuspensão do pellet celular em 50µl em água Milli-Q livre de nuclease seguido por 10 minutos de aquecimento à 95°C. Em seguida, o produto é resfriado a -80°C durante 20 minutos. Após esse período, o produto é descongelado a temperatura ambiente e adicionado 30µg de Proteinase K seguido de uma incubação a 56°C por 30 minutos e, logo após, a temperatura é elevada para 95°C e a incubação é realizada por mais 10 minutos. Então o produto é centrifugado a 16.000g por 1 minuto seguido pela coleta do sobrenadante para obtenção do DNA. O armazenamento indicado é a -20°C.

#### 2.5. Extração de RNA e síntese de cDNA

As extrações de RNA das linhagens celulares foram realizadas utilizando o reagente TRIzol® (Invitrogen) de acordo com indicado pelo fabricante. De modo geral, a lise celular foi realizada pela adição de Trizol seguida pela adição de clorofórmio para separação da fase aquosa. Depois utiliza-se isopropanol absoluto para precipitação do RNA e a lavagem do pellet é realizada com etanol gelado a 75%. Por fim, utiliza-se água livre de RNase para ressuspender o pellet seguida pela incubação a 55-60°C por 15 minutos. Armazenamento a -80°C. A integridade do RNA foi analisada através da visualização das bandas do RNA ribossomal 28S e 18S por eletroforese em gel de agarose a 1%. Apenas as amostras de RNA que estavam íntegras seguiram para a fase da síntese de cDNA.

Para síntese de cDNA, o RNA foi primeiro tratado com o kit DNase I (Sigma-Aldrich), seguindo as instruções do fabricante. Para a reação da transcrição reversa foi utilizado o kit "High-Capacity cDNA Reverse Transcription" (AppliedBiosystems) conforme o indicado pelo fabricante.

## 2.6. PCR convencional e PCR quantitativa (qPCR)

O DNA genômico foi utilizado a 50ng/μl para realização da PCR convencional utilizando o kit "GoTaq® G2 Hot Start Taq Polymerase" seguindo as seguintes condições no termociclador: 94°C a 2 minutos, 94°C por 30 segundos, 55°C por 30 segundos, 72°C por 30 segundos e 72°C por 5 minutos, com repetição de 30 ciclos. Amostras foram corridas em gel de agarose a 1% a 100 volts. A PCR convencional foi utilizada para visualizar a presença da Cas9 (Exxtend) e da Beta actina (Invitrogen) na linhagem celular de escolha para o prosseguimento do estudo (Tabela 1).

Para qPCR o kit "PowerUp™ SYBR™ Green Master Mix" e o sistema StepOne Plus™ (Applied Biosystems) foram utilizados. O cDNA obtido das linhagens celulares REH, RS4;11, Nalm-6 e 697 foram dispostos em triplicata em placa de 96 poços para avaliação da expressão do *SMYD5* (Invitrogen) (Tabela 1), sendo 1μl de amostra para 9μl do Master Mix por poço. Testes para curvas de diluição com eficiência entre 90-110% de acordo com MIQE (*Minimum Information for Publication of Quantitative Real-Time PCR*) (28) foram realizados para cada amostra em uma razão de 2 (1:2 – 1:64). A diluição de melhor eficiência utilizada para os ensaios de expressão gênica foi de 1:8.

**Tabela 1:** Sequência de primers.

Gene	Sequência	Orientação	Aplicação
Cas9	CTGCAGACAGTGAAGGTGGT	Forward	PCR
	TCGGCCTTGGTCAGATTGTC	Reverse	PCR
Beta actina	GGACTTCGAGCAAGAGATGG	Forward	PCR
	AGCACTGTGTTGGCGTACAG	Reverse	PCR
<i>SMYD5</i>	CCCCTTGCATCCTCTCAATA	Forward	qPCR
	TCTCCGCAGAAGTTCCAGTT	Reverse	qPCR

### 2.6.1. Análise de dados da qPCR

Para a quantificação relativa da expressão do mRNA do *SMYD5* entre as quatro linhagens celulares, o método comparativo de quantificação relativa ( $\Delta\Delta C_T$ ) foi utilizado. Resumidamente, a média dos valores de  $C_T$  das amostras analisadas do gene de interesse foram normalizadas (valor de  $C_T$  de cada amostra subtraído pela média do valor de  $C_T$  do gene endógeno), achando o valor de  $\Delta C_T$  e em seguida, os valores de  $\Delta\Delta C_T$  são calculados. Como nesse estudo foram utilizados dois genes

endógenos, a média geométrica dos genes *TBP* e *B2M* foram calculadas. Para calcular a expressão gênica, o  $\Delta\Delta C_T$  de cada replicata do gene alvo foi dividida pela média geométrica dos genes endógenos. Por fim, a média e o desvio padrão dos valores foram calculados e organizados utilizando o GraphPad prism 8.0.

## 2.7. Produção de lentivirus e Transfecção de linhagem celular

Para obtenção de expressão estável da endonuclease Cas9 na linhagem selecionada, utilizou-se o sistema de infecção por lentivirus com plasmídeos previamente estabelecidos no CancerLab, baseando-se em protocolos também previamente estabelecidos (29). Resumidamente, para confecção do lentivírus, 3 plasmídeos foram utilizados:

1) um plasmídeo que contenha o gene da endonuclease Cas9, que irá se integrar no genoma do hospedeiro.

2) plasmídeo pCMV Delta R8.2, responsável por codificar todos os componentes para o empacotamento do lentivírus;

3) plasmídeo pCMV VSVG, que codifica a proteína do envelope viral.

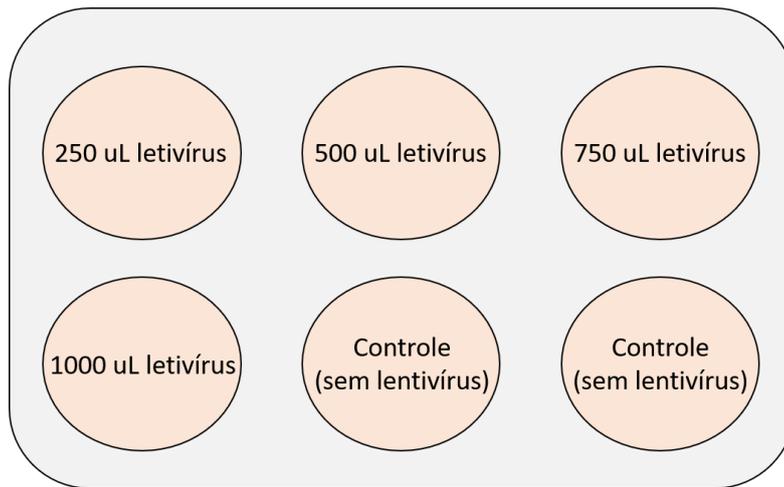
Há então a junção dos três plasmídeos em proporção equimolar ao meio Opti-MEM (Invitrogen) e lipofectamina 2000 (Invitrogen). Após 20 minutos de incubação, a mistura é adicionada a célula HEK293T e o período de incubação é de 4 a 6 horas. Após esse período, o meio de transfecção é removido e o meio DMEM completo é adicionado a célula. O meio de cultivo é então coletado em 24h, 48h e 72h depois e a concentração do vírus é realizada de acordo com o protocolo "4×Lentivirus Concentrator Solution". Em seguida, o produto é armazenado a  $-80^{\circ}\text{C}$  até o uso.

### 2.7.1. Transdução lentiviral

O concentrado lentiviral é então adicionado a célula de interesse, em um processo chamado de transdução para geração de células estáveis que expressam cas9. Em resumo, esse processo ocorre da seguinte forma:

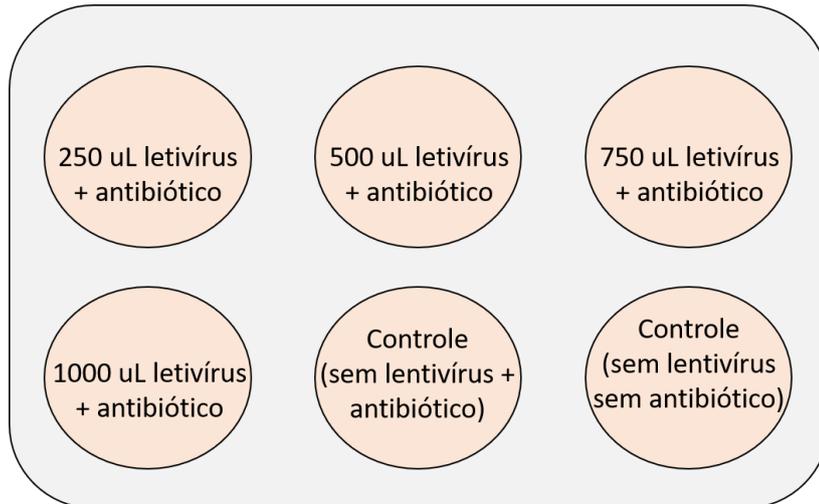
- Diferentes diluições do lentivírus são preparadas em meio de cultivo completo mais polibreno ( $16,0\mu\text{g}/\text{mL}$ ) para determinar a diluição de melhor eficiência (ver imagem 2).

- 1mL de cada diluição é adicionada em cada poço de uma placa de 6. Recomenda-se deixar dois poços para controle sem a presença do lentivírus, do qual um poço receberá antibiótico para posterior seleção celular.
- Adicionar 300.000 células em suspensão em 1mL de meio completo em cada poço. Aqui, a concentração de polibreno ficará em 8ug/mL.
- Incubar entre 48-72 horas a 37°C com CO<sub>2</sub> a 5%.



**Figura 3:** Representação da configuração da placa no processo de transdução.

- Depois do período de incubação, há a troca do meio completo com a adição do antibiótico adequado. Nessa etapa, o processo de seleção de células estáveis é iniciado (imagem 3).
- Após a etapa anterior, é necessário a observação diária da placa para garantir que há morte celular nos poços controle. A troca do meio deve ser realizada a cada 2-3 dias para manter a dose correta do antibiótico.
- Esse processo de seleção dura o tempo em que as células do poço controle + antibióticos levam para morrer por completo. Aqui, a concentração utilizada foi de 1µg/mL de puromicina (P7255, Sigma).



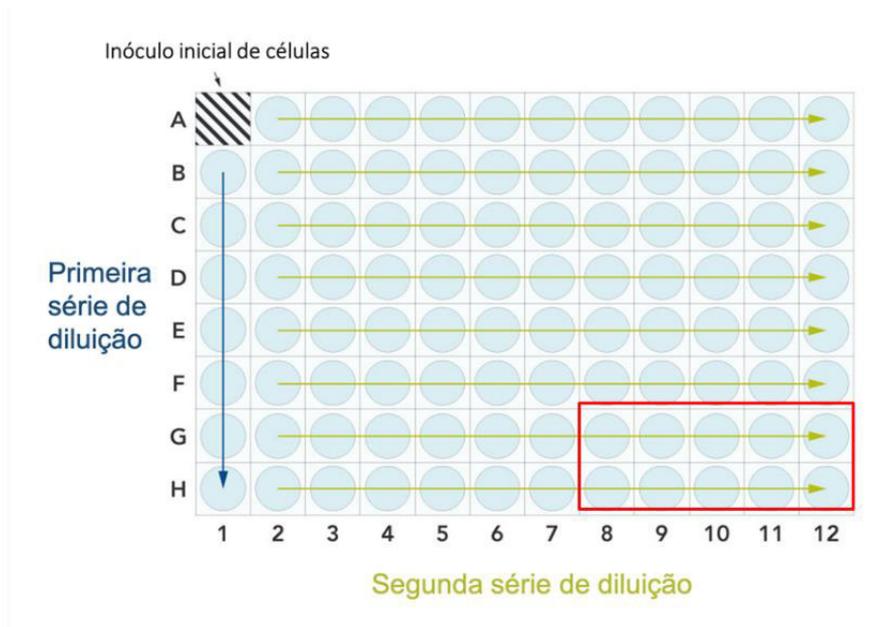
**Figura 4:** Representação da configuração da placa no processo de seleção de células transduzidas.

- Houve então a expansão das populações transduzidas e a população com melhor expressão foi escolhida para separação de clones de célula única.

#### 2.7.2. Separação clonal

Para obtenção de uma população de células monoclonais, utilizou-se o protocolo alternativo de diluição seriada, e a verificação da expressão estável da Cas9 foi feita por PCR convencional e posteriormente confirmada por western blot. Para isso, foi necessário a inoculação de  $4 \times 10^3$  células no poço A1 de uma placa de 96 poços, o que permite uma diluição em série primeiro na vertical seguida por outra diluição seriada na horizontal em proporção de 1:2. Com isso, estima-se que consiga cerca de 10 clones por cada placa de 96 poços (Figura 4).

Recomenda-se o uso de meio condicionado, que é o meio completo de uma placa que esteja de 50-60% de confluência, que possui fatores de crescimento que vão estimular a proliferação das células individuais. Ao coletar esse meio, recomenda-se que seja filtrado através de um filtro PES de 0,45  $\mu\text{m}$ . O meio condicionado pode ser diluído em meio novo completo na proporção de 1:1 a 20% de SFB.



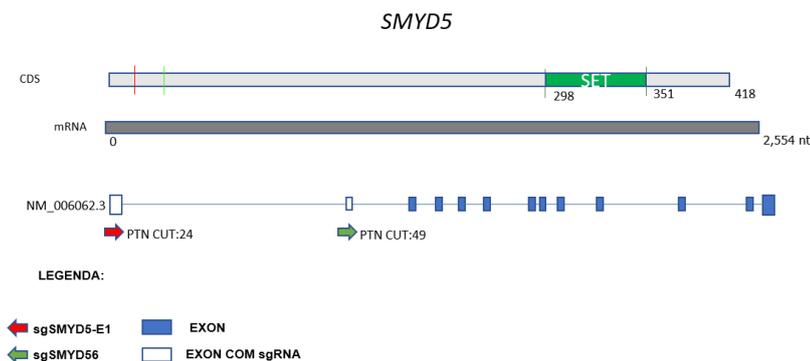
**Figura 5.** Representação da configuração da placa de 96 poços para o método de diluição de matriz seriada. Marcado em vermelho estão os poços que, teoricamente, encontram-se células únicas.

### 2.7.3. Transfecção celular utilizando sgRNA

Para realização do nocaute gênico, sgRNAs (single-guide RNAs, em inglês) foram desenhados com o auxílio da plataforma CHOPCHOP (<https://chopchop.cbu.uib.no/>) (30), e adquiridos comercialmente. A transfecção dos guias de RNA foi feita por lipofecção diretamente nas linhagens Cas9 estáveis de interesse, e o protocolo alternativo de diluição seriada para obtenção de células únicas foi novamente realizado conforme descrito acima para obter clones monoclonais com K.O do *SMYD5*. O processo de transfecção de guia foi realizada de acordo com o protocolo “Transfecção de células Cas9-estáveis com TrueGuide™ Synthetic gRNA” utilizando reagente de transfecção Lipofectamine™ RNAiMAX™” (Invitrogen), com a opção para placa de 6 poços. Foram confeccionadas 2 guias para o K.O gênico, sendo uma direcionada para o éxon 1 e outra para o éxon 2 (Tabela 2; Figura 5).

**Tabela 2.** Sequência de sgRNA utilizadas para transfecção em células Cas9 estáveis para K.O gênico.

Linhagem celular	Guia	Código da guia	Éxon	Fita	GC%
REH-cas9	GCTCACGAAACGGACTTCCA CGG	SMYD5-E1	1	Negativa	55
	CAGGGGCCGTTCTACGAAGA TGG	SMYD5-E2	2	Negativa	60



**Figura 6.** Representação do SMYD5 com as guias confeccionadas para knockout gênico direcionadas para o Éxon 1 e Éxon 2 na linhagem celular de leucemia linfóide aguda.

## 2.8. Sequenciamento de Sanger

Para confirmar o nocaute gênico, é realizada a extração de DNA seguindo o protocolo alternativo descrito acima. As amostras são então diluídas à 5ng/μL para realização da PCR convencional. Os primers utilizados para amplificação e sequenciamento da região de interesse encontram-se na (tabela 3). Para algumas amostras houve a purificação do gel usando o kit "GenElute gel extraction" (Sigma) segundo as instruções do fabricante. O produto amplificado é então enviado para sequenciamento e o resultado é analisado através do programa Geneious Prime.

**Tabela 3:** Sequencia de primers utilizados para amplificação e sequenciamento da região de interesse do *SMYD5* para confirmação de knockout gênico. Produtos produzidos pela Exxtend.

Sequência	Éxon	Orientação	Aplicação
AGGCACCGATGTAAAAGCAC	1	Forward	PCR
CCAACCATACACACAGCCAC	1	Reverse	PCR
GGTTAAGGGTCATAAGGCGG	1	Forward	Sequenciamento
GAAATGCGCACGGAGACTC	1	Reverse	Sequenciamento
CAGCCTGCTATAAGCCTTGG	2	Forward	PCR
GCCTAAGTCCTGCTGTGCTC	2	Reverse	PCR
AAGAGTGAGGCATGTGGGAG	2	Forward	Sequenciamento
CACATCTGAAGAGCCTCCCA	2	Reverse	Sequenciamento

## 2.9. Western Blot

Para realização do western blot, proteínas foram extraídas das linhagens celulares utilizando RIPA contendo inibidor de protease, e a concentração proteica utilizada foi de 35ug por poço.

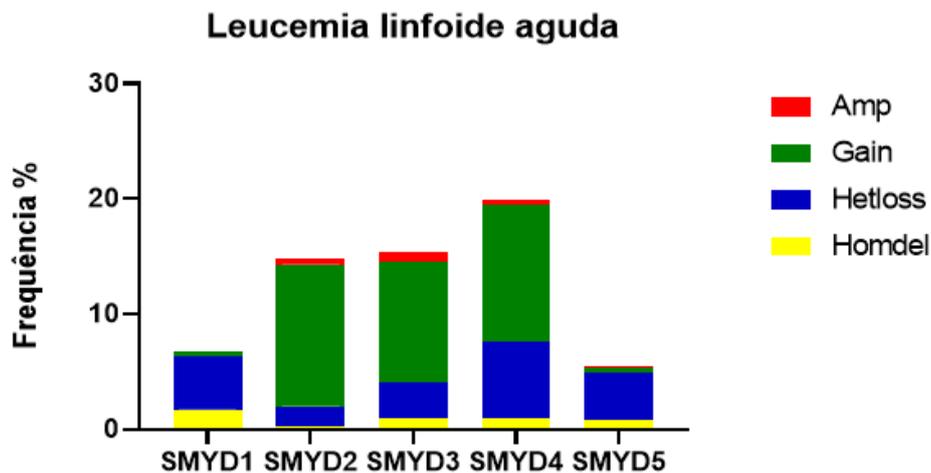
De modo geral, os lisados de proteínas são aplicados em gel de poliacrilamida (gel concentrador a 4% e separador a 10%) onde ocorre a separação das proteínas de acordo com seu tamanho por eletroforese em gel com voltagem fixa a 110V. Em seguida, foi realizada a transferência das proteínas do gel para membrana de PVDF (fluoreto de polivinilideno) de 0,22 µm utilizando o método “wet” a 24V overnight. Após o processo de transferência, a membrana é bloqueada com uma solução bloqueadora (PBS-T + leite em pó desnatado a 5%). Em seguida, houve a incubação por 1 hora a temperatura ambiente sob agitação com o anticorpo primário “monoclonal Anti-Flag® M2 (Merk, F1804-200G) e “Anti-Beta actin” (Abcam, AC-15) padronizados em diluição de 1:1.000 cada.

O passo seguinte foi a lavagem com PBS-T e incubação com o anti-mouse diluído em solução de bloqueio a 2% a temperatura ambiente por 1 hora, em diluição de 1:10.000 para a membrana marcada com anti-flag e 1:15.000 para membrana marcada com anti-B-actin. A visualização da membrana foi realizada com fotodocumentador Amersham™ Imager 600 (GE life science) utilizando o “Luminata™ Forte Western HRP Substrate” (Millipore) para detecção em quimioluminescência. A quantificação das imagens digitalizadas foi realizada com software ImageJ.

### 3. RESULTADOS

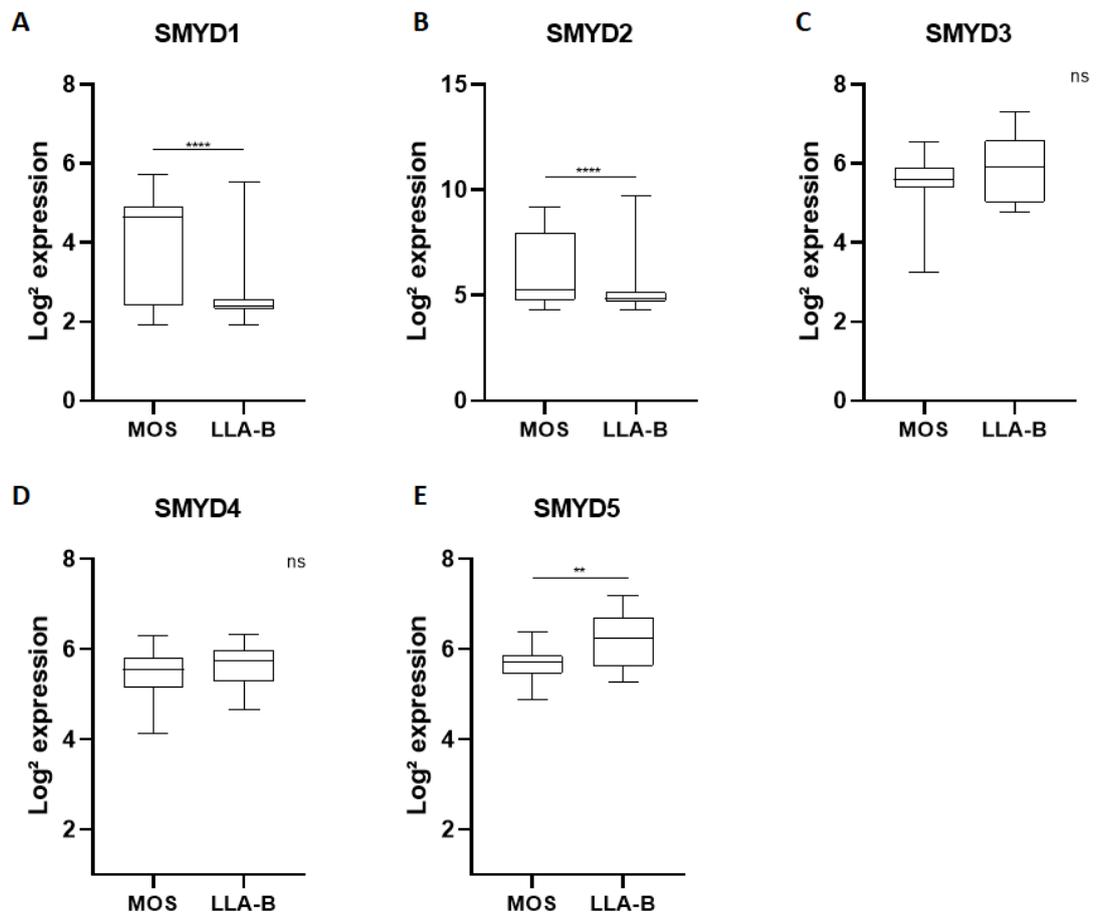
#### 3.1. Avaliação do número de cópias alteradas e expressão do mRNA

Os dados referentes ao número de cópias alteradas dos genes da família *SMYD* foram igualmente obtidos através do cBioPortal e analisados usando o algoritmo GISTIC como descrito no capítulo anterior (12). A alteração de maior frequência na LLA observada no *SMYD1* e *SMYD5* foi a perda heterozigótica (4,6% e 4,1%, respectivamente), enquanto que o *SMYD2*, *SMYD3* e *SMYD4* tiveram o ganho de baixo nível (12,3%, 10,5% e 11,9%, respectivamente) como a alteração mais frequente (Figura 6, Tabela Anexo I).



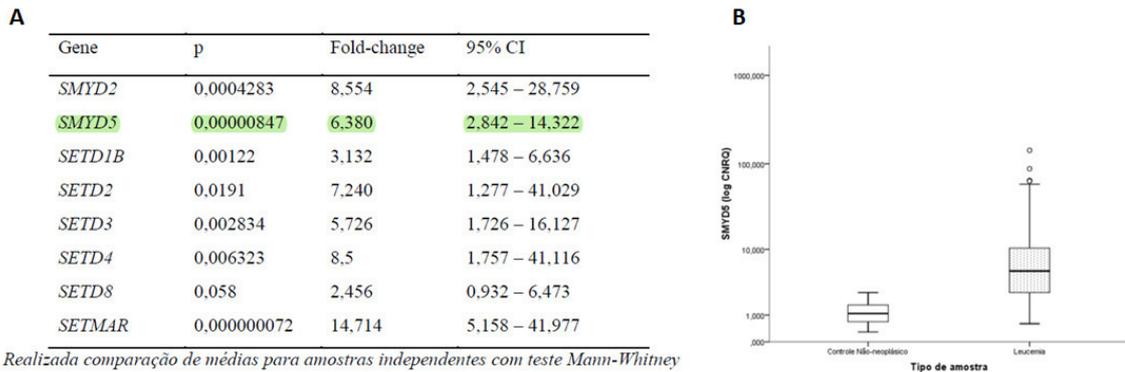
**Figura 7.** Frequência de alterações no número de cópias na família *SMYD* na Leucemia Linfoide Aguda de células B. n = 763. (Amp – amplificação de alto nível; Gain – ganho de baixo nível; Hetloss – deleção heterozigótica; Homdel – deleção homozigótica).

Em relação a expressão do mRNA, a base de dados BloodSpot foi utilizada para analisar o nível de expressão dos cinco membros da família *SMYD* comparando-os com dados da expressão dos mesmos advindos de amostras de medula óssea saudável. Houve diferença estatisticamente significativa nos genes *SMYD1* e *SMYD2* que apresentaram-se com baixa expressos quando comparados a amostras da medula óssea saudável ( $p = <0,0001$ ) (Figura 7 A e B). O *SMYD5*, por sua vez, apresentou alta expressão quando comparado a amostras saudáveis da medula óssea ( $p = 0,0022$ ) (Figura 7 E). Os *SMYD3* e *SMYD4* não tiveram diferença estatisticamente significativas (Figura 7 C e D).



**Figura 8:** Expressão do mRNA da família *SMYD* em medula óssea saudável e leucemia linfóide aguda de células B. **(A)** *SMYD1*: MOS n= 131; LLA-B n= 72. **(B)** *SMYD2*: MOS n= 167; LLA-B n= 107. **(C)** *SMYD3*: MOS n= 73; LLA-B n= 14. **(D)** *SMYD4*: MOS n= 73; ALL-B n= 13. **(E)** *SMYD5*: MOS n= 73; LLA-B n= 13. (MOS – Medula óssea saudável). Teste de Mann Whitney (\*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001).

Com o achado de Sakamoto et.al (2014) do qual descreveu a superexpressão do *SMYD5* em LLA infantil (Figura 8), juntamente com o resultado aqui apresentado (Figura 7), o *SMYD5* foi escolhido como alvo de investigação para realização de knockout gênico e averiguação de seu envolvimento com o crescimento e proliferação celular.



**Figura 9:** Resultados retirados da tese “Papel de metiltransferases de proteínas no desenvolvimento e prognóstico de leucemias linfóides agudas da infância” de Luís Sakamoto (2014). **(A)** Comparação dos níveis de expressão de metiltransferases de lisinas em amostras de leucemias e medula óssea não-neoplásica, com destaque para o *SMYD5*. **(B)** Comparação do nível de expressão do *SMYD5* em amostras de controle não neoplásico e amostras leucêmicas.

### 3.1.1. Correlação entre CNA e mRNA do *SMYD5*

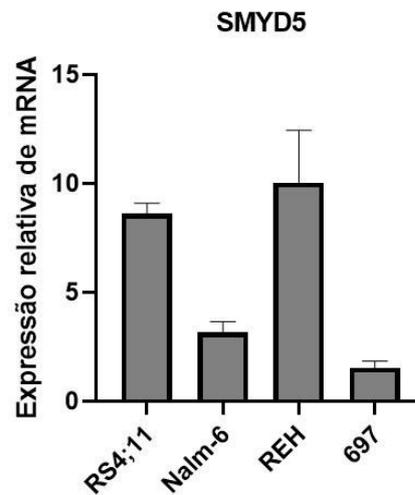
Em seguida, dados obtidos através do *cBioPortal* foram analisados para saber se havia correlação entre o CNA e a expressão do mRNA do *SMYD5*, gene de interesse do presente estudo nesse capítulo. Contudo, constatou-se que não há correlação entre o CNA e o mRNA (Tabela 4).

**Tabela 4.** Correlação entre CNA e expressão do mRNA do *SMYD5*

Gene	Pacientes (n)	Pearson	Valor de p	Spearman	Valor de p
<i>SMYD5</i>	11	0,0364	0,915	0,000	0,0909

### 3.2. Nível de expressão do *SMYD5* em linhagens celulares de LLA-B

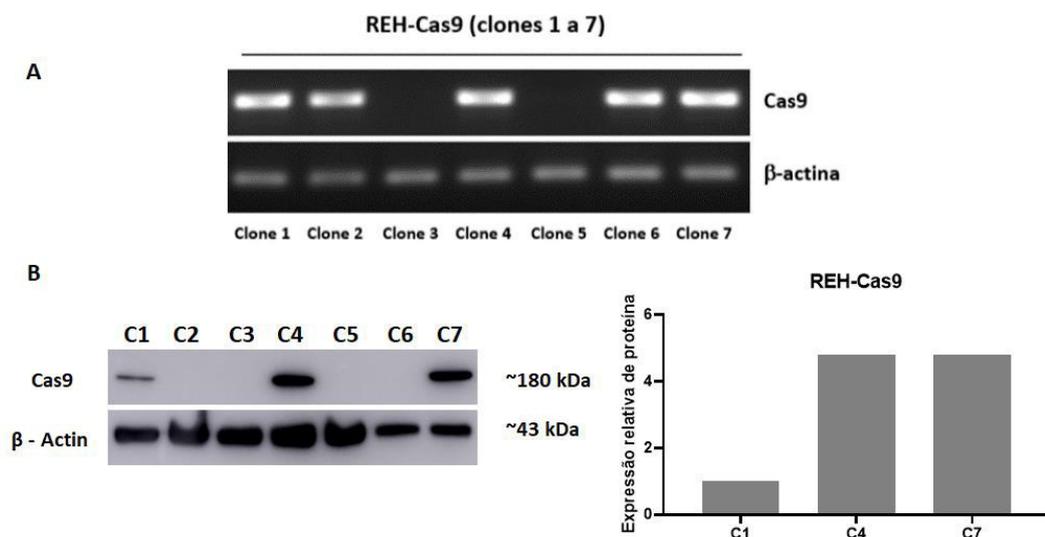
Para análise *in vitro*, o *SMYD5* foi selecionado e, dessa forma, seu nível de expressão foi avaliado em quatro linhagens celulares de leucemia linfóide aguda do tipo B (RS4;11, REH, Nalm-6 e 697) através de qPCR. Com isso, a linhagem celular REH apresentou uma maior expressão do *SMYD5* em relação às outras linhagens celulares, sendo então selecionada para os testes subsequentes (Figura 8).



**Figura 10.** Análise de qPCR para avaliar a expressão do *SMYD5* em linhagens celulares de leucemia linfóide aguda de células.

### 3.3. Seleção de clone celular com expressão estável de Cas-9

Em seguida, a linhagem REH foi submetida a transdução lentiviral para obtenção de clones de células únicas com expressão de Cas-9 estável (Figura 9 A). Dessa forma, sete clones foram obtidos e testados a nível proteico. Com isso, o clone 4 foi o escolhido para o knockout do *SMYD5* por apresentar uma boa expressão proteica da Cas-9 entre os clones avaliados (Figura 9 B).



**Figura 11. (A)** PCR para confirmar expressão de Cas9 e Beta actina nos clones 1 ao 7 da linhagem REH-Cas9. **(B)** Western Blot para avaliar nível proteico da Cas-9 nos clones 1 ao 7 obtidos da REH-Cas9. Gráfico gerado através do ImageJ.

#### 4. DISCUSSÃO

Este capítulo analisou as modificações moleculares como alteração do número de cópias e a expressão de mRNA na leucemia linfóide aguda de célula B presentes na família *SMYD*, tendo o *SMYD5* como alvo principal do estudo, por meio do banco de dados cBioportal e BloodSpot. As análises dessas alterações lançaram luz sobre o potencial uso desses genes como biomarcadores.

Resultados da expressão da família *SMYD* em LLA-B obtidos pelo BloodSpot mostraram que a expressão do mRNA do *SMYD1*, *SMYD2* e *SMYD5* tiveram diferença significativa quando comparados a amostras advindas de medula óssea saudáveis (MOS), enquanto que a expressão do mRNA do *SMYD3* e *SMYD4* não tiveram diferença estatisticamente significativa comparadas a MOS. Ao comparar as alterações do CNA com a expressão do mRNA, apenas o *SMYD1* entrou em concordância, mas isso pode acontecer pelo fato da alteração do CNA não necessariamente afetar a expressão do mRNA (31).

O *SMYD1* mostrou-se com baixa expressão na LLA-B quando comparado a amostras de MOS, e teve como alteração mais frequente no CNA a perda heterozigótica, o que entra em concordância. Porém, não há até o momento estudos que correlacione o *SMYD1* com a LLA-B. Todavia, Peveling-Oberhag et.al (2015) demonstrou que o *SMYD1* continha mutações no linfoma não-Hodking de célula B, não sendo possível caracterizar se o *SMYD1* agia como oncogene ou supressor tumoral (19). Dessa forma, há um indício de que o *SMYD1* possua algum envolvimento em células B que merece uma melhor elucidação em estudos futuros.

Em relação ao *SMYD2*, o achado desse estudo apresentou uma redução de sua expressão gênica quando comparado a MOS, não estando de acordo com achados em estudos anteriores dos quais apresentaram o *SMYD2* superexpresso em LLA-B, resultando em um pior prognóstico com redução do tempo de sobrevivência dos pacientes (20,21). Portanto, apesar de Brown et.al (2020) também corroborar os resultados citados acima, foi demonstrado que em camundongos o *knockout* do *SMYD2* em células tronco hematopoiéticas gerou alterações na hematopoiese pelo fato de sua perda levar a modulações na via WNT/  $\beta$ -catenina (32). Contudo, é inegável a participação do *SMYD2* em LLA-B e seu potencial como alvo terapêutico.

A expressão do mRNA do *SMYD3* e *SMYD4* não tiveram diferença estatisticamente significativas quando comparadas a MOS e, até o momento, não há estudos que os correlacionem com a LLA. Porém, na leucemia linfóide crônica, que é uma malignidade que afeta células B, há estudos que mostram que a alta expressão do STAT3 leva ao aumento da expressão do *SMYD3*, promovendo a progressão da LLC, podendo ser usado como biomarcador para prognóstico de agressividade, contribuindo com o mau prognóstico da doença (22,23,33). Portanto, apesar de não haver associação com a LLA-B, há associação da alteração do *SMYD3* em células B que levam a progressão da LLC indicando, dessa forma, que uma melhor investigação sobre o *SMYD3* na LLA-B pode ser promissora.

No presente estudo, o *SMYD5* mostrou-se superexpresso, o que condiz com os achados de Sakamoto et.al (2014) que descreveu um aumento de expressão do *SMYD5* em LLA infantil quando comparado a amostras de medula óssea não neoplásicas. Nesse caso, não foi encontrada associação com o tempo de sobrevida, porém, esse aumento de expressão em amostras de LLA merece uma melhor compreensão (24). Apesar da pouca informação encontrada na literatura sobre o *SMYD5* não só em relação a malignidades hematopoiéticas, mas no processo carcinogênico em geral, o mesmo já foi descrito como superexpresso no câncer gástrico e em processos metastáticos do câncer de mama para o pulmão (34,35).

Um estudo realizado por Fujii et.al (2016) em zebrafish mostrou que o *SMYD5* possui funções fundamentais na hematopoiese e também na mielopoiese por interagir com a nucleofosmina 1 (NPM1) (25,36), que por sua vez é encontrada mutada em 25%-35% dos adultos diagnosticados com LMA (37). Outro estudo revelou que o *SMYD5* forma um complexo juntamente com NCoR (co-repressor do receptor nuclear), responsável pela marca da trimetilação na lisina 20 da histona 4 (H4K20me3) na região promotora de um grupo de genes responsáveis pela resposta pró-inflamatória, mantendo-os reprimidos por restringir o receptor toll-like 4 (TLR-4) (38). Portanto, uma redução da expressão do *SMYD5* poderá desencadear uma ativação constante de genes inflamatórios que podem contribuir com o processo da carcinogênese, já que o processo inflamatório é um dos *hallmarks* do câncer (39).

Em relação a LLA, foi sugerido por Guidez et.al (2000) que o NCoR possui uma função repressora no transcrito da proteína TEL (conhecida por ETV6), e provavelmente na oncoproteína de fusão TEL-AML1 (40). É possível especular que o

*SMYD5* participe da repressão da TEL ao formar um complexo com o NCoR. Entretanto, uma investigação mais aprofundada se faz necessária para uma compreensão mais completa de seu papel no rearranjo TEL-AML1, que é a alteração mais comum em crianças com LLA-B. (9). Já a baixa expressão do TLR4 também foi descrita em PBMC de crianças com LLA-B quando comparado ao controle, que sugere uma possível explicação para redução da resposta imune antitumoral na LLA (41).

Durante o processo de diferenciação de células tronco embrionárias, a depleção do *SMYD5* ocasiona aberrações cromossômicas, característica comum em leucemias, e a transformação celular (9,42). Esse processo pode levar ao desenvolvimento de células cancerígenas por apresentar redução das marcas epigenéticas H4K20me3 e H3K9me3, marcas essas que são relacionadas a diversos tipos tumorais (42). A H4K20me3 é uma marca envolvida com a formação da heterocromatina que conseqüentemente regula e reprime a transcrição, sendo comum em ocorrências oncológicas que reduz de acordo com a agressividade tumoral. Já a H3K9me3 auxilia na estabilidade genômica e sua perda relaciona-se a células cancerígenas (42–44). Portanto, nesta etapa de diferenciação embrionária, é possível que o *SMYD5* tenha papel importante ao evitar a redução da H4K20me3 e H3K9me3, aberrações cromossômicas e a transformação de células cancerígenas. Porém, a expressão constitutiva desregulado do *SMYD5* pode manter também a repressão de genes pró-inflamatórios impedindo a ligação do TLR4 e afetar, dessa forma, a resposta imunológica contra células cancerígenas. Essa dualidade de possíveis mecanismos de ação no processo da tumorigênese salienta a necessidade de futuras investigações sobre o envolvimento do *SMYD5* nessas duas vias apresentadas, e enfatiza a importância da sua desregulação no contexto da leucemia linfóide aguda de células B.

Dessa forma, através dos resultados prévios encontrados nesse estudo, juntamente com os achados sobre o *SMYD5* na hematopoiese e oncogênese discutidos acima, um modelo celular de LLA-B que expressa a proteína Cas-9 de forma estável foi construído afim de averiguar, em estudos futuros, a função do *SMYD5* nos processos de proliferação, viabilidade e crescimento celular através do seu knockout utilizando a tecnologia de Crispr-Cas9.

## 5. CONCLUSÃO

Ao avaliar as alterações presentes na família SMYD de metiltransferases, com foco maior no *SMYD5*, para uma melhor compreensão do envolvimento desses genes na LLA-B afim de determinar potenciais alvos moleculares mais eficazes para o monitoramento e tratamento da LLA-B, foi determinado:

- *SMYD1* e *SMYD2* tiveram baixa expressão em pacientes com LLA-B.
- O *SMYD5*, por sua vez, apresentou-se com a expressão elevada em amostras de LLA-B.
- Os *SMYD1*, *SMYD2* e *SMYD5* podem servir como potenciais biomarcadores moleculares para pacientes com LLA-B.

## 6. PERSPECTIVAS

Os resultados obtidos neste trabalho e os dados disponíveis na literatura evidenciam a importância de análise mais aprofundada do papel do *SMYD5* na leucemogênese de LLA. As ferramentas que construímos para o knockout do *SMYD5* em linhagens celulares de LLA-B via CRISPR/Cas9, permitirão a realização de estudos futuros para elucidar os efeitos do *SMYD5* no crescimento, proliferação e viabilidade celular. Assim como permitirão explorar a interação do *SMYD5* com genes pró-inflamatórios e sua relação com o desenvolvimento e progressão da LLA-B.

## 7. BIBLIOGRAFIA

1. Metabolic underpinnings of leukemia pathology and treatment - PubMed [Internet]. [cited 2024 Jan 28]. Available from: <https://pubmed.ncbi.nlm.nih.gov/32721091/>
2. Anjos AR dos, Alvares-Silva M, Borelli P. Matriz Extracelular e Leucemia. *Rev Bras Hematol Hemoter.* 2000 Dec;22:404–12.
3. Minciacchi VR, Kumar R, Krause DS. Chronic Myeloid Leukemia: A Model Disease of the Past, Present and Future. *Cells.* 2021 Jan 10;10(1):117.
4. Hamerschlak N. Leucemia: fatores prognósticos e genética. *J Pediatr (Rio J).* 2008 Aug;84:S52–7.
5. Stubbins RJ, Francis A, Kuchenbauer F, Sanford D. Management of Acute Myeloid Leukemia: A Review for General Practitioners in Oncology. *Curr Oncol.* 2022 Aug 30;29(9):6245–59.
6. Hallek M, Shanafelt TD, Eichhorst B. Chronic lymphocytic leukaemia. *Lancet.* 2018 Apr 14;391(10129):1524–37.
7. Chiorazzi N, Chen SS, Rai KR. Chronic Lymphocytic Leukemia. *Cold Spring Harb Perspect Med.* 2021 Feb 1;11(2):a035220.
8. Malard F, Mohty M. Acute lymphoblastic leukaemia. *Lancet.* 2020 Apr 4;395(10230):1146–62.
9. Lejman M, Chałupnik A, Chilimoniuk Z, Dobosz M. Genetic Biomarkers and Their Clinical Implications in B-Cell Acute Lymphoblastic Leukemia in Children. *Int J Mol Sci.* 2022 Mar 2;23(5):2755.
10. Forghieri F, Luppi M, Potenza L. Philadelphia chromosome-positive Acute Lymphoblastic Leukemia. *Hematology.* 2015 Dec;20(10):618–9.
11. Duffield AS, Mullighan CG, Borowitz MJ. International Consensus Classification of acute lymphoblastic leukemia/lymphoma. *Virchows Arch.* 2023 Jan;482(1):11–26.
12. Song J, Liu Y, Chen Q, Yang J, Jiang Z, Zhang H, et al. Expression patterns and the prognostic value of the SMYD family members in human breast carcinoma using integrative bioinformatics analysis. *Oncol Lett.* 2019 Apr;17(4):3851–61.
13. Liu D, Wang X, Shi E, Wang L, Nie M, Li L, et al. Comprehensive Analysis of the Value of SMYD Family Members in the Prognosis and Immune Infiltration of Malignant Digestive System Tumors. *Front Genet.* 2021;12:699910.
14. Zhang Y, Yan L, Yao W, Chen K, Xu H, Ye Z. Integrated Analysis of Genetic Abnormalities of the Histone Lysine Methyltransferases in Prostate Cancer. *Med Sci Monit.* 2019 Jan 7;25:193–239.
15. Reynoird N, Mazur PK, Stellfeld T, Flores NM, Lofgren SM, Carlson SM, et al. Coordination of stress signals by the lysine methyltransferase SMYD2 promotes pancreatic cancer. *Genes Dev.* 2016 Apr 1;30(7):772–85.

16. Hamamoto R, Furukawa Y, Morita M, Imura Y, Silva FP, Li M, et al. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol.* 2004 Aug;6(8):731–40.
17. Hu L, Zhu YT, Qi C, Zhu YJ. Identification of Smyd4 as a potential tumor suppressor gene involved in breast cancer development. *Cancer Res.* 2009 May 1;69(9):4067–72.
18. SMYD5 acts as a potential biomarker for hepatocellular carcinoma - PubMed [Internet]. [cited 2023 Dec 28]. Available from: <https://pubmed.ncbi.nlm.nih.gov/35218722/>
19. Peveling-Oberhag J, Wolters F, Döring C, Walter D, Sellmann L, Scholtysik R, et al. Whole exome sequencing of microdissected splenic marginal zone lymphoma: a study to discover novel tumor-specific mutations. *BMC Cancer.* 2015 Oct 24;15:773.
20. Sakamoto LHT, Andrade RV de, Felipe MSS, Motoyama AB, Pittella Silva F. SMYD2 is highly expressed in pediatric acute lymphoblastic leukemia and constitutes a bad prognostic factor. *Leukemia Research.* 2014 Apr 1;38(4):496–502.
21. Zhang P, Ruan J, Weng W, Tang Y. Overexpression of SET and MYND domain-containing protein 2 (SMYD2) is associated with poor prognosis in pediatric B lineage acute lymphoblastic leukemia. *Leuk Lymphoma.* 2020 Feb;61(2):437–44.
22. Ma W, Zhang Y, Qi Y, Guo S. STAT3 promotes chronic lymphocytic leukemia progression through upregulating SMYD3 expression. *Arch Med Sci.* 2019 Sep;15(5):1163–75.
23. Lin F, Wu D, Fang D, Chen Y, Zhou H, Ou C. STAT3-induced SMYD3 transcription enhances chronic lymphocytic leukemia cell growth in vitro and in vivo. *Inflamm Res.* 2019 Sep;68(9):739–49.
24. SAKAMOTO, Luis Henrique Toshihiro. Papel de metiltransferases de proteínas no desenvolvimento e prognóstico de leucemias linfóides agudas da infância. Universidade de Brasília; 2014.
25. Fujii T, Tsunesumi SI, Sagara H, Munakata M, Hisaki Y, Sekiya T, et al. Smyd5 plays pivotal roles in both primitive and definitive hematopoiesis during zebrafish embryogenesis. *Sci Rep.* 2016 Jul 5;6:29157.
26. Reyes DA, Sarría VMS, Salazar-Viedma M, D'Afonseca V. Histone Methyltransferases Useful in Gastric Cancer Research. *Cancer Inform.* 2021;20:11769351211039862.
27. Bagger FO, Kinalis S, Rapin N. BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Research.* 2019 Jan 8;47(D1):D881–5.
28. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 2009 Apr;55(4):611–22.
29. Tandon N, Thakkar KN, LaGory EL, Liu Y, Giaccia AJ. Generation of Stable Expression Mammalian Cell Lines Using Lentivirus. *Bio Protoc.* 2018 Nov 5;8(21):e3073.

30. Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research*. 2019 Jul 2;47(W1):W171–4.
31. Bhattacharya A, Bense RD, Urzúa-Traslaviña CG, de Vries EGE, van Vugt MATM, Fehrmann RSN. Transcriptional effects of copy number alterations in a large set of human cancers. *Nat Commun*. 2020 Feb 5;11(1):715.
32. Brown MA, Edwards MA, Alshirahi I, Geng H, Dekker JD, Tucker HO. The lysine methyltransferase SMYD2 is required for normal lymphocyte development and survival of hematopoietic leukemias. *Genes Immun*. 2020 Feb;21(2):119–30.
33. Oliveira-Santos W, Rabello DA, Lucena-Araujo AR, de Oliveira FM, Rego EM, Pittella Silva F, et al. Residual expression of SMYD2 and SMYD3 is associated with the acquisition of complex karyotype in chronic lymphocytic leukemia. *Tumour Biol*. 2016 Jul;37(7):9473–81.
34. Reyes DA, Sarría VMS, Salazar-Viedma M, D'Afonseca V. Histone Methyltransferases Useful in Gastric Cancer Research. *Cancer Inform*. 2021;20:11769351211039862.
35. Gao H, Chakraborty G, Lee-Lim AP, Mavrakis KJ, Wendel HG, Giancotti FG. Forward genetic screens in mice uncover mediators and suppressors of metastatic reactivation. *Proc Natl Acad Sci U S A*. 2014 Nov 18;111(46):16532–7.
36. Abu-Farha M, Lanouette S, Elisma F, Tremblay V, Butson J, Figeys D, et al. Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. *J Mol Cell Biol*. 2011 Oct;3(5):301–8.
37. Hindley A, Catherwood MA, McMullin MF, Mills KI. Significance of NPM1 Gene Mutations in AML. *Int J Mol Sci*. 2021 Sep 17;22(18):10040.
38. Stender JD, Pascual G, Liu W, Kaikkonen MU, Do K, Spann NJ, et al. Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. *Mol Cell*. 2012 Oct 12;48(1):28–38.
39. Hallmarks of Cancer: The Next Generation - ScienceDirect [Internet]. [cited 2023 Jun 27]. Available from: <https://www.sciencedirect.com/science/article/pii/S0092867411001279>
40. Guidez F, Petrie K, Ford AM, Lu H, Bennett CA, MacGregor A, et al. Recruitment of the nuclear receptor corepressor N-CoR by the TEL moiety of the childhood leukemia-associated TEL-AML1 oncoprotein. *Blood*. 2000 Oct 1;96(7):2557–61.
41. Sánchez-Cuaxospa M, Contreras-Ramos A, Pérez-Figueroa E, Medina-Sansón A, Jiménez-Hernández E, Torres-Nava JR, et al. Low expression of Toll-like receptors in peripheral blood mononuclear cells of pediatric patients with acute lymphoblastic leukemia. *Int J Oncol*. 2016 Aug;49(2):675–81.
42. Kidder BL, He R, Wangsa D, Padilla-Nash HM, Bernardo MM, Sheng S, et al. SMYD5 Controls Heterochromatin and Chromosome Integrity during Embryonic Stem Cell Differentiation. *Cancer Res*. 2017 Dec 1;77(23):6729–45.
43. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet*. 2005 Apr;37(4):391–400.

44. Feinberg AP, Koldobskiy MA, Göndör A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet.* 2016 May;17(5):284–99.

## 8. ANEXO I

**Tabela Anexo I:** Frequência total do número de cópias alteradas referentes a Leucemia Linfóide Aguda.

Gene	Amplificação	Ganho de baixo nível	Perda heterozigótica	Perda homozigótica
<i>SMYD1</i>	0,0%	0,5%	4,6%	1,7%
<i>SMYD2</i>	0,5%	12,3%	1,7%	0,3%
<i>SMYD3</i>	0,8%	10,5%	3,1%	0,9%
<i>SMYD4</i>	0,4%	11,9%	6,7%	0,9%
<i>SMYD5</i>	0,1%	0,5%	4,1%	0,8%



Article

# Genetic Alterations of *SMYD4* in Solid Tumors Using Integrative Multi-Platform Analysis

Brunna Letícia Olivera Santana, Mariana Braccialli de Loyola , Ana Cristina Moura Gualberto and Fabio Pittella-Silva \*

Laboratory of Molecular Pathology of Cancer, Faculty of Healthy Sciences, University of Brasília, Federal District, Brasília 70910-900, Brazil; brunna.los@hotmail.com (B.L.O.S.); marianabraccialli@outlook.com (M.B.d.L.); ana.gualberto@unb.br (A.C.M.G.)

\* Correspondence: pittella@unb.br

**Abstract:** *SMYD4* is a member of the *SMYD* family that has lysine methyltransferase function. Little is known about the roles of *SMYD4* in cancer. The aim of this study is to investigate genetic alterations in the *SMYD4* gene across the most prevalent solid tumors and determine its potential as a biomarker. We performed an integrative multi-platform analysis of the most common mutations, copy number alterations (CNAs), and mRNA expression levels of the *SMYD* family genes using cohorts available at the Cancer Genome Atlas (TCGA), cBioPortal, and the Catalogue of Somatic Mutations in Cancer (COSMIC). *SMYD* genes displayed a lower frequency of mutations across the studied tumors, with none of the *SMYD4* mutations detected demonstrating sufficient discriminatory power to serve as a biomarker. In terms of CNAs, *SMYD4* consistently exhibited heterozygous loss and downregulation across all tumors evaluated. Moreover, *SMYD4* showed low expression in tumor samples compared to normal samples, except for stomach adenocarcinoma. *SMYD4* demonstrated a frequent negative correlation with other members of the *SMYD* family and a positive correlation between CNAs and mRNA expression. Additionally, patients with low *SMYD4* expression in STAD and LUAD tumors exhibited significantly poorer overall survival. *SMYD4* demonstrated its role as a tumor suppressor in the majority of tumors evaluated. The consistent downregulation of *SMYD4*, coupled with its association with cancer progression, underscores its potential usefulness as a biomarker.

**Keywords:** *SMYD* family; *SMYD4*; solid tumors; in silico analysis; protein methyltransferase; epigenetics



**Citation:** Olivera Santana, B.L.; de Loyola, M.B.; Gualberto, A.C.M.; Pittella-Silva, F. Genetic Alterations of *SMYD4* in Solid Tumors Using Integrative Multi-Platform Analysis. *Int. J. Mol. Sci.* **2024**, *25*, 6097. <https://doi.org/10.3390/ijms25116097>

Academic Editor: Saverio Candido

Received: 30 April 2024

Revised: 21 May 2024

Accepted: 22 May 2024

Published: 31 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Protein methyltransferases (PMTs) are a class of epigenetic modifiers essential for regulating a wide range of biological processes. Their misregulation can contribute to the development of various physiopathological conditions, including cancers. PMTs participate in crucial epigenetic processes by catalyzing the transfer of methyl groups to lysine or arginine residues. Molecular alterations in PMTs due to copy number alterations (CNA) or aberrant mRNA expression can lead to a series of dysfunctional events affecting several physiological conditions. The dysregulation of methylation processes has been linked to various pathological conditions, including cancer development, progression, and increased tumor aggressiveness, as well as metastatic events [1,2].

The *SMYD* (SET and MYND domain-containing proteins) family is a subgroup of protein methyltransferases that consists of five members (*SMYD1-5*). These proteins contain the SET domain, responsible for lysine methylation. The SET domain is divided into two segments by the MYND domain (Myeloid, Nervy, and DEAF-1), which is a unique characteristic of the *SMYD* family that facilitates protein-protein interactions [3]. While *SMYD4* has been less studied in carcinogenesis compared to other members of the *SMYD* family, it has been identified as a potential tumor suppressor due to its observed downregulation in breast cancer [4–6]. High expression of *SMYD4* has been implicated in the signaling pathways of cancer stem cells (CSCs), contributing to tumor cell transformation [7]. However, a

comprehensive understanding of the molecular alterations in *SMYD4* within the context of a specific cancer remains largely unexplored.

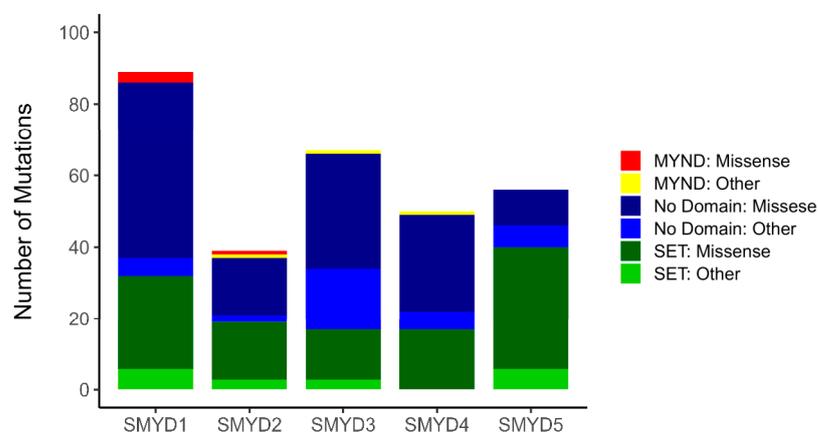
According to the World Health Organization, the four most prevalent tumors among males in 2022, excluding non-melanoma skin cancer, were lung cancer (15.3%), prostate cancer (14.2%), colorectal cancer (10.4%), and stomach cancer (6.1%). In females, the leading types were breast cancer (23.8%), lung cancer (9.4%), colorectal cancer (8.9%), and cervix cancer (6.8%) [8]. Despite the heterogeneity of these cancer types, classical molecular signatures, such as mutations or alterations in key cell cycle regulators, like TP53, are often described [9]. Whether these cancer types also share common alterations in epigenetic genes is still unknown.

In this study, we conducted an integrative multi-platform analysis using publicly available databases to investigate the prevalent molecular alterations in the genes of the *SMYD* family, specifically in *SMYD4*, across eight of the most common tumor types. We analyzed data from breast invasive carcinoma (BRCA), prostate adenocarcinoma (PRAD), colorectal carcinoma (CRC), and stomach adenocarcinoma (STAD), as well as the most common subtypes of lung cancer, namely lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). Additionally, we examined uterine corpus endometrial carcinoma (UCEC) and uterine carcinosarcoma (UCS) cohorts from the Cancer Genome Atlas (TCGA) and cBioPortal. We aimed to uncover the potential prognostic values associated with these alterations.

## 2. Results

### 2.1. Mutation Profile of *SMYD* Genes

We initially analyzed mutations in each *SMYD* family gene across the eight tumor types under examination. We used cohorts available at the cBio Cancer Genomics portal as well as at the Catalogue of Somatic Mutations in Cancer (COSMIC). The mutations in *SMYD* genes were examined in a total of 9696 patients encompassing all cancer types included in the analysis. Missense mutations were individually analyzed, as they were the most prevalent among the identified mutations, while other non-missense mutation types, such as frameshift, insertions, or deletions, were grouped and analyzed collectively. A total of 240 missense and 64 non-missense mutations were identified across the *SMYD* genes. The majority of missense mutations were located in non-conserved regions (129), followed by the SET domain (107) and the MYND domain (4). Specifically, *SMYD4* exhibited a total of 50 mutations, with 32 located in non-conserved regions, 17 in the SET domain and 1 in the MYND domain (Figure 1).



**Figure 1.** Total somatic mutations separated by conserved domains and non-conserved regions. Missense mutations in the SET domain are represented in dark green, and other mutations are noted in light green. Missense mutations in the MYND domain are represented in red, and other mutations are noted in yellow. Missense mutations in non-specific regions are represented in dark blue, and other mutations are noted in light blue.

It was observed that the mutation frequencies within the *SMYD* family were relatively low, with certain genes exhibiting mutation frequencies higher than 2%. Specifically, *SMYD5* exhibited a mutation frequency of 2.3% in colorectal cancer, while *SMYD1* demonstrated a mutation frequency of 2.3% in both lung adenocarcinoma and lung squamous cell carcinoma. In endometrial carcinoma, the mutation frequencies were 3.3% for *SMYD1*, 2.1% for *SMYD2*, 2.1% for *SMYD3*, 2.1% for *SMYD4*, and 2.4% for *SMYD5*. Moreover, in uterine carcinosarcoma, the mutation frequencies were 5.1% for *SMYD1* and 2.5% for *SMYD4* (Table 1). These findings indicate that uterine tumors exhibited the highest mutation frequencies among the *SMYD* genes within the tumors evaluated, with particular emphasis on the mutation frequency of *SMYD1* in uterine carcinosarcoma.

**Table 1.** Mutation frequency in each *SMYD* family member among the eight tumor types.

Tumor	<i>SMYD1</i>	<i>SMYD2</i>	<i>SMYD3</i>	<i>SMYD4</i>	<i>SMYD5</i>
CRC	1.4%	0.9%	1.0%	0.5%	2.3%
STAD	1.2%	0.9%	1.9%	0.8%	1.1%
BRCA	0.3%	0.2%	0.4%	0.4%	0.3%
PRAD	0.2%	0.1%	0.2%	0.1%	0.1%
Lung Cancer					
LUAD	2.3%	0.8%	0.7%	0.4%	0.5%
LUSC	2.3%	0.8%	1.2%	0.8%	0.4%
Uterine Cancer					
UCEC	3.3%	2.1%	2.1%	2.1%	2.4%
UCS	5.1%	0.0%	0.0%	2.5%	1.3%

However, no specific mutations across the *SMYD* genes were found to be frequently recurrent among patients with the same tumor type or within different tumors.

## 2.2. Copy Number Alterations in the *SMYD* Family Genes

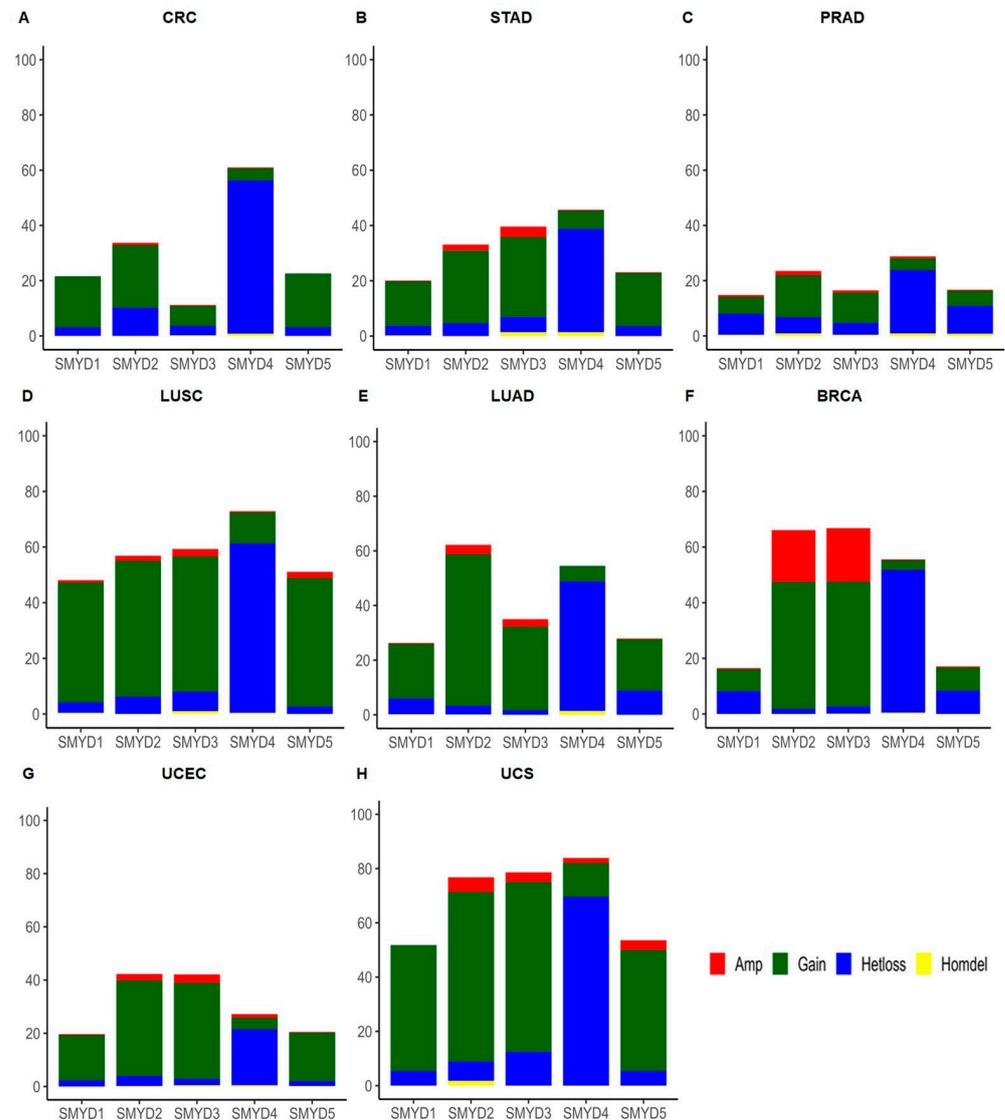
Copy number alteration (CNA) refers to changes in DNA copy numbers occurring at specific locations in the genome. These alterations can lead to the activation of oncogenes or the suppression of tumor suppressor genes. Understanding the functionality of this mechanism is crucial for advancing the development of potential therapeutic and diagnostic markers [10]. We obtained CNA data for the *SMYD* genes through cBioPortal and analyzed these data using the GISTIC algorithm across all evaluated tumors [11].

*SMYD4* exhibited a uniform alteration pattern, marked by a high incidence of heterozygous loss across all examined tumor types. These included colorectal adenocarcinoma (55.5%), stomach adenocarcinoma (37.4%), prostate adenocarcinoma (22.9%), lung squamous cell carcinoma (60.9%), lung adenocarcinoma (47.4%), breast carcinoma (51.3%), endometrial carcinoma (21%), and uterine carcinosarcoma (69.6%) (Figure 2A–H). Importantly, six out of eight tumors had heterozygous loss as the predominant CNA with a frequency of more than 30%. No other gene in the *SMYD* family exhibited a similar, consistently high level of heterozygous loss across the different tumor types.

In the case of *SMYD1*, the majority of alterations were observed as low-level gains in colorectal adenocarcinoma (18.5%), stomach adenocarcinoma (16.1%), lung squamous cell carcinoma (43.1%), lung adenocarcinoma (20%), endometrial carcinoma (17.1%), and uterine carcinosarcoma (46.4%). Prostate adenocarcinoma (7.6%) and breast carcinoma (8.2%) exhibited heterozygous loss as the most frequent alteration, although low-level gains exhibited frequencies of 6.2% and 7.9%, respectively (Figure 2A–H).

In all the tumors evaluated, low-level gains emerged as the most common alteration observed in *SMYD2* and *SMYD3*. Notably, in the case of breast carcinoma, the frequency of high-level amplification accounted for 18.6% and 19.2% of the alterations in *SMYD2* and *SMYD3*, respectively (Figure 2A–H). Finally, low-level gain was the most frequent alteration in *SMYD5* in colorectal adenocarcinoma (19.5%), stomach adenocarcinoma (19.2%),

squamous cell lung cancer (46.3%), lung adenocarcinoma (18.7%), endometrial carcinoma (18.2%), and uterine carcinosarcoma (44.6%) tumors. Prostate adenocarcinoma (10.1%) was characterized by heterozygous loss. Breast carcinoma had similar distributions of low-level gain and heterozygous loss, with frequencies of 8.3% and 8.4%, respectively (Figure 2A–H).

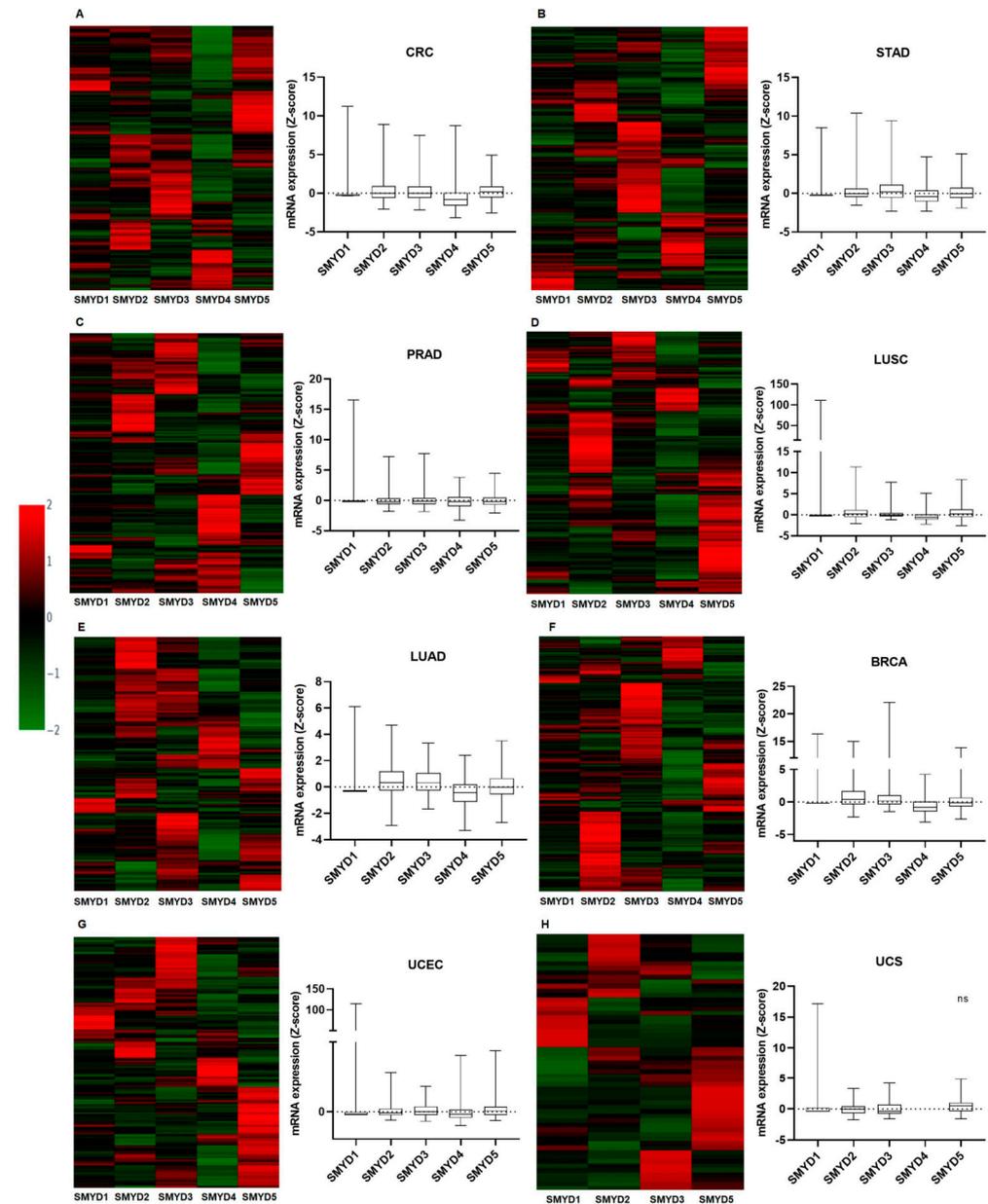


**Figure 2.** Frequency of copy number alterations in *SMYD* family genes. (A) Colorectal adenocarcinoma:  $n = 616$  for *SMYD1*, 2, 4, and 5;  $n = 1943$  for *SMYD3*. (B) Stomach adenocarcinoma:  $n = 589$ . (C) Prostate adenocarcinoma:  $n = 2848$  for *SMYD1*, 2, and 5;  $n = 3890$  for *SMYD3*;  $n = 1835$  for *SMYD4*. (D) Lung squamous cell carcinoma:  $n = 501$ . (E) Lung adenocarcinoma:  $n = 1109$  for *SMYD1*, 2, and 4;  $n = 1973$  for *SMYD3*;  $n = 1158$  for *SMYD5*. (F) Invasive breast cancer:  $n = 3469$ . (G) Endometrial cancer:  $n = 620$ . (H) Uterine carcinosarcoma:  $n = 56$ . (Amp—high level amplification; Gain—low-level gain; Hetloss—heterozygous deletion; Homdel—homozygous deletion).

### 2.3. mRNA Expression Profile

mRNA expression was assessed using Z-scores relative to diploid sample data obtained for each cancer type from cBioPortal. Heatmaps and boxplots were used to assess the expression levels of the *SMYD* genes in all eight cancer types. In line with previous results observed for CNA, *SMYD4* also exhibited a uniform expression pattern in all tumors evaluated. In all tumors cohorts, *SMYD4* followed a pattern of downregulation (Figure 3A–H). This result corroborates the CNA findings, which indicated a high frequency of heterozygous

loss (Figure 2A–H). These observations are consistent with previous studies suggesting a potential tumor suppressor function for *SMYD4* [4–6]. Among other *SMYD* family genes, *SMYD4* was the only one with such a consistent downregulation pattern across the different tumor types.



**Figure 3.** mRNA expression of *SMYD* family genes in solid tumors. (A) Colorectal cancer (n = 382). (B) Stomach adenocarcinoma (n = 415). (C) Prostate adenocarcinoma (*SMYD1*, 4, and 5: n = 498; *SMYD2*: n = 486; *SMYD3*: n = 493). (D) Lung squamous cell cancer (n = 501). (E) Lung adenocarcinoma (n = 169). (F) Invasive breast cancer (1100). (G) Uterine corpus endometrial carcinoma (n = 177). (H) Uterine carcinosarcoma (n = 57). Low expression was defined as a value below the 50th percentile, and high expression was defined as a value above the 50th percentile. *SMYD4* expression data for UCS were not available.

*SMYD1* was downregulated in CRC, BRCA, UCEC, and UCS tumors, while its expression was upregulated in STAD, PRAD, LUSC, and LUAD. A similar variation in mRNA expression levels was observed for *SMYD2*. While its expression was increased in the majority of CRC, STAD, LUSC, LUAD, and BRCA tumors, it was downregulated in patients

with PRAD, UCEC and UCS tumors. *SMYD3* showed upregulated expression levels in CRC, STAD, LUAD, BRCA, and UCEC, while it was downregulated in the majority of PRAD and UCS patients. *SMYD3* remained at basal levels in LUSC (Figure 3A–H). *SMYD5* was upregulated in the majority of CRC, STAD, LUSC, LUAD, UCEC, and UCS tumors. In contrast, it was downregulated in PRAD and BRCA patients (Figure 3A–H).

With the exception of *SMYD4*, which exhibited a consistent downregulated pattern, all the other members of the *SMYD* family showed a distinct pattern of expression depending on the tumor.

#### 2.4. Correlation Analysis of *SMYD4* Expression with Other *SMYD* Genes

The correlation of *SMYD4* expression with other *SMYD* family genes was assessed using Pearson correlation analysis. The analysis revealed weak negative correlations between *SMYD4* and *SMYD3* ( $r = -0.136$ ,  $p = 0.008$ ) and between *SMYD4* and *SMYD5* ( $r = -0.184$ ,  $p = 3.015 \times 10^{-4}$ ) in CRC (Figure 4A). In STAD, there was no significant correlation between *SMYD4* and the other *SMYD*s (Figure 4B). In PRAD, Pearson's correlation also showed poor negative correlations between *SMYD4* and *SMYD2* ( $r = -0.141$ ,  $p = 0.002$ ), *SMYD4* and *SMYD3* ( $r = -0.249$ ,  $p = 1.761 \times 10^{-8}$ ), and *SMYD4* and *SMYD5* ( $r = -0.268$ ,  $p = 1.309 \times 10^{-9}$ ) (Figure 4C).

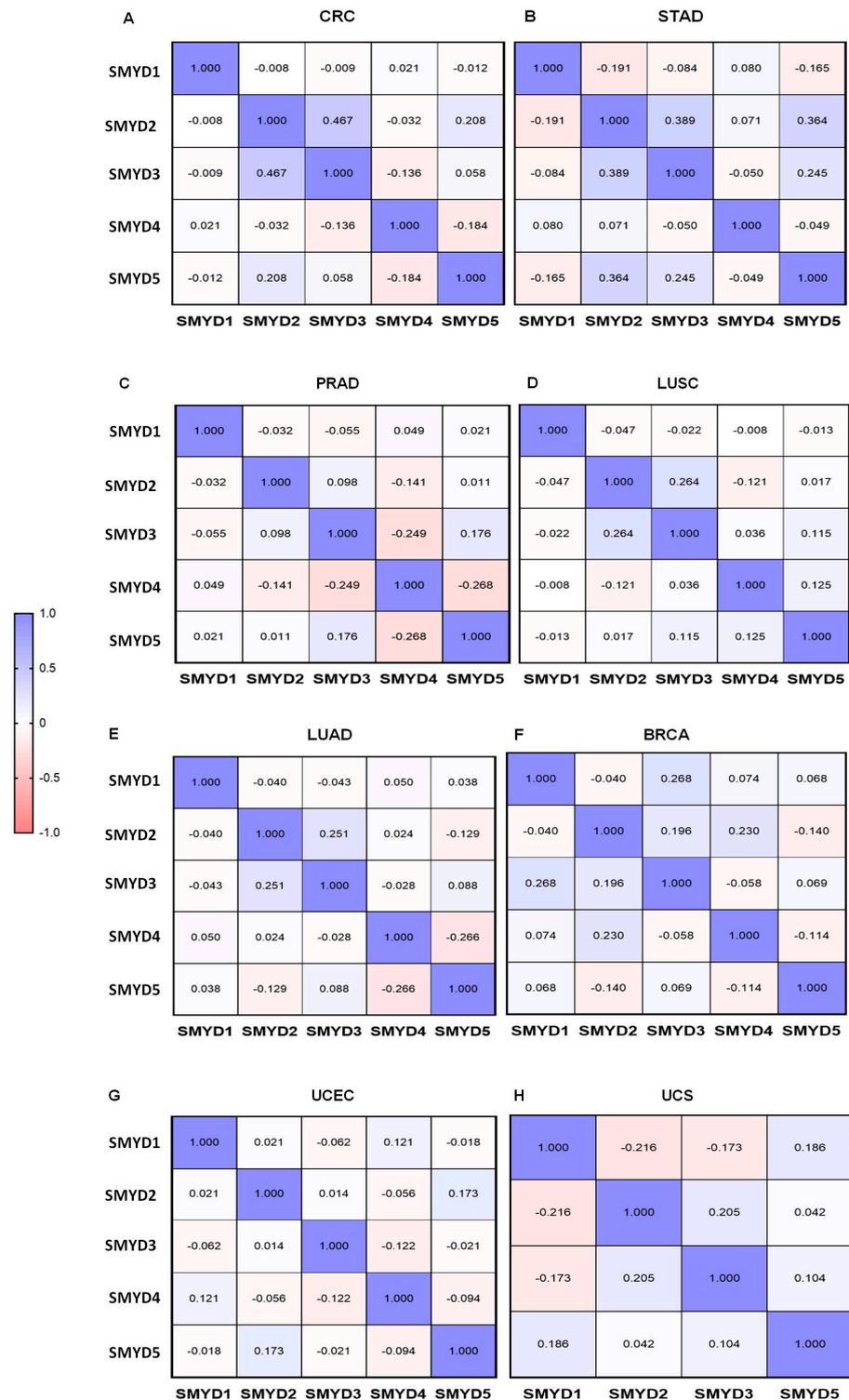
In LUSC, a slightly positive correlation was observed between *SMYD4* and *SMYD5* ( $r = 0.125$ ,  $p = 0.005$ ). A poor negative correlation was observed between *SMYD2* and *SMYD4* ( $r = -0.121$ ,  $p = 0.007$ ) (Figure 4D). In LUAD, a weak negative correlation was also observed between *SMYD4* and *SMYD5* ( $r = -0.266$ ,  $p = 4.666 \times 10^{-4}$ ) (Figure 4E).

There was also a minimal positive correlation between *SMYD1* and *SMYD4* ( $r = 0.074$ ,  $p = 6.045 \times 10^{-5}$ ) and a low negative correlation between *SMYD3* and *SMYD4* ( $r = -0.058$ ,  $p = 4.967 \times 10^{-5}$ ) and between *SMYD4* and *SMYD5* ( $r = -0.114$ ,  $p = 0.002$ ) in BRCA (Figure 4F). In UCEC, a poor positive correlation was observed between *SMYD2* and *SMYD4* ( $r = -0.056$ ,  $p = 0.002$ ) (Figure 4G). In UCS, there was no significant correlation between *SMYD4* and the other members of the *SMYD* family (Figure 4H).

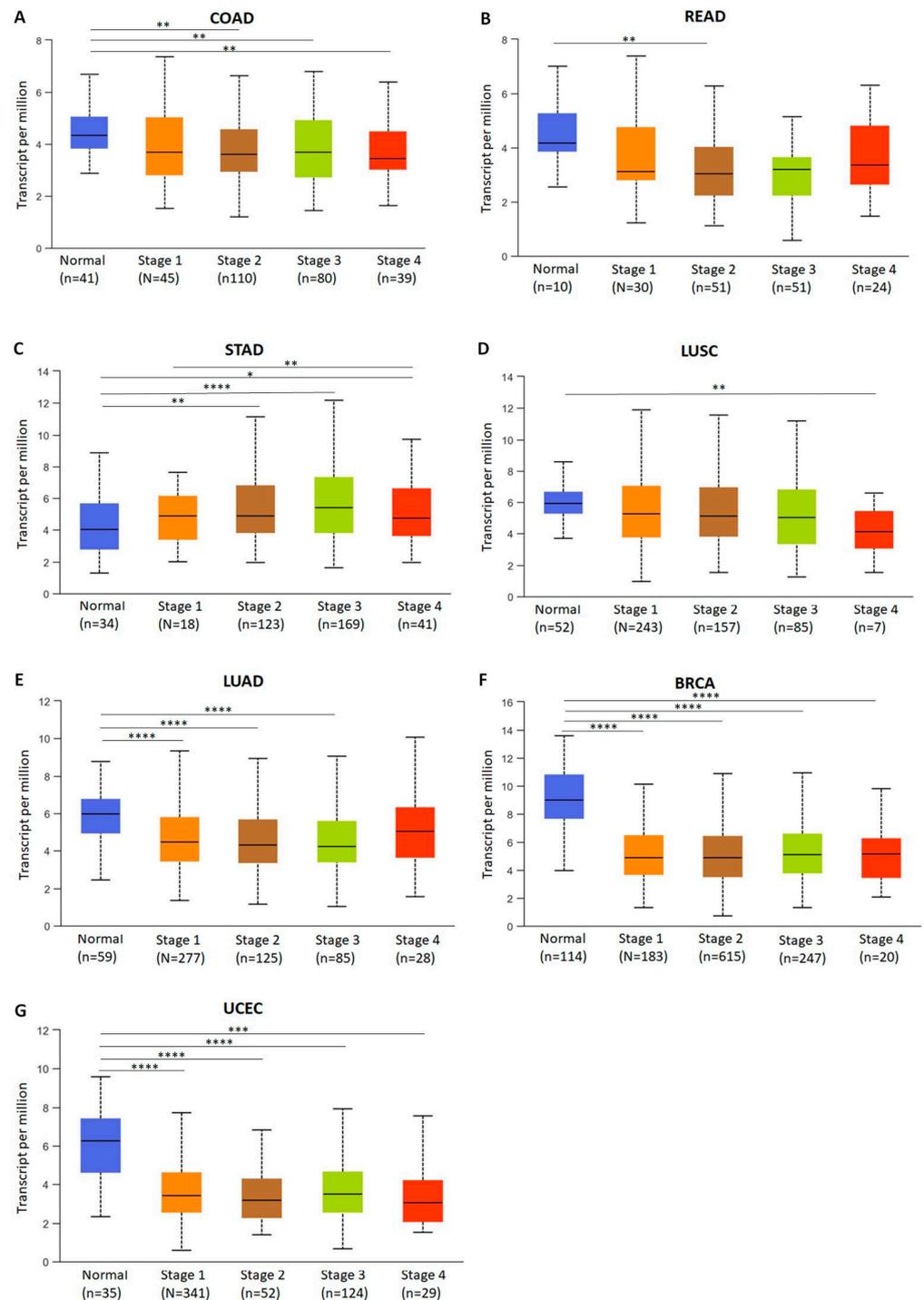
#### 2.5. Comparison of *SMYD4* Expression between Normal Samples and Tumor Samples at Different Tumor Stages

*SMYD4* exhibited a negative regulation pattern based on mRNA analysis in all the evaluated tumors. To assess the expression of *SMYD4* in normal samples and tumor samples at different disease stages, we profiled its mRNA expression using the UALCAN database. In this analysis, colon and rectal tumors were evaluated separately, and it was observed that *SMYD4* expression in colon adenocarcinoma was significantly lower in stage 2, 3, and 4 tumors ( $p = 1.04 \times 10^{-3}$ ,  $5.55 \times 10^{-3}$ , and  $1.21 \times 10^{-2}$ , respectively) when compared to normal samples. In rectal adenocarcinoma, significantly low expression of *SMYD4* was only observed in stage 2 tumor samples ( $p = 1.86 \times 10^{-2}$ ) when compared to normal sample (Figure 5A,B).

The expression of *SMYD4* in STAD was noteworthy for the fact that it differed from the other tumors in that its expression was significantly higher in stages 2, 3, and 4 when compared to the normal sample ( $p = 3.83 \times 10^{-3}$ ,  $2.28 \times 10^{-5}$ , and  $3.37 \times 10^{-2}$ , respectively). Stage 3 had a statistically significant higher expression of *SMYD4* when compared to stage 1 ( $p = 1.94 \times 10^{-2}$ ) (Figure 5C). In LUSC, *SMYD4* expression was significantly lower in stage 4 compared to normal samples ( $p = 1.55 \times 10^{-3}$ ) (Figure 5D). In LUAD, patients with stage 1, 2, and 3 tumors presented low expression when compared to normal samples ( $p = 1.19 \times 10^{-6}$ ,  $p = 7.68 \times 10^{-8}$ , and  $p = 3.51 \times 10^{-5}$ , respectively) (Figure 5E). In BRCA, *SMYD4* expression was also significantly lower in all tumor stages in comparison with normal samples ( $p < 1 \times 10^{-12}$ ,  $p = 1.11 \times 10^{-16}$ ,  $p < 1 \times 10^{-12}$ , and  $p = 1.96 \times 10^{-12}$  for stages 1, 2, 3, and 4, respectively) (Figure 5F). In UCEC, low expression of *SMYD4* was observed in stages 1, 2, 3, and 4 ( $p = 3.26 \times 10^{-11}$ ,  $7.71 \times 10^{-10}$ ,  $1.30 \times 10^{-9}$ , and  $7.80 \times 10^{-4}$ , respectively) when comparing them with normal samples (Figure 5G).



**Figure 4.** Correlation of mRNA expression of SMYD family genes in solid tumors. (A) Colorectal cancer (n = 382). (B) Stomach adenocarcinoma (n = 415). (C) Prostate adenocarcinoma (SMYD1, 4, and 5: n = 498; SMYD2: n = 486; SMYD3: n = 493). (D) Lung squamous cell cancer (n = 501). (E) Lung adenocarcinoma (n = 169). (F) Invasive breast cancer (1100). (G) Uterine corpus endometrial carcinoma (n = 177). (H) Uterine carcinosarcoma (n = 57). Pearson correlation test was used for comparison of expression levels among the five genes in each tumor evaluated ( $p = 0.05$ ). Red represents a negative correlation ( $r$  values close to  $-1$ ). Blue represents a positive correlation ( $r$  values close to  $+1$ ). White represents independent variables with values close to 0.



**Figure 5.** SMYD4 expression between normal and tumor stages samples. (A) Colon adenocarcinoma (COAD). (B) Rectum adenocarcinoma (READ). (C) Stomach adenocarcinoma (STAD). (D) Lung squamous cell carcinoma (LUSC). (E) Lung adenocarcinoma (LUAD). (F) Invasive breast cancer (BRCA). (G) Uterine corpus endometrial carcinoma (UCEC). No data were available for normal samples comparison to UCS. (\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ; \*\*\*\*  $p < 0.0001$  with statistical significance).

## 2.6. Correlation Analyses between SMYD4 CNAs and mRNA Expression

A correlation analysis was conducted to ascertain the relationship between SMYD4 copy number alterations (CNAs) and mRNA expression across various tumor types. These included colorectal cancer (CRC), gastric adenocarcinoma (STAD), lung squamous cell carcinoma (LUSC), breast invasive carcinoma (BRCA), and uterine corpus endometrial

carcinoma (UCEC). We found a statistically significant positive correlation for all tumors evaluated in this study, indicating that CNAs influenced mRNA expression in these tumors. Specifically, CRC, STAD, LUSC, and BRCA had a Pearson correlation coefficient greater than 0.500. LUSC and UCEC had Spearman correlation coefficient greater than 0.500 (Table 2).

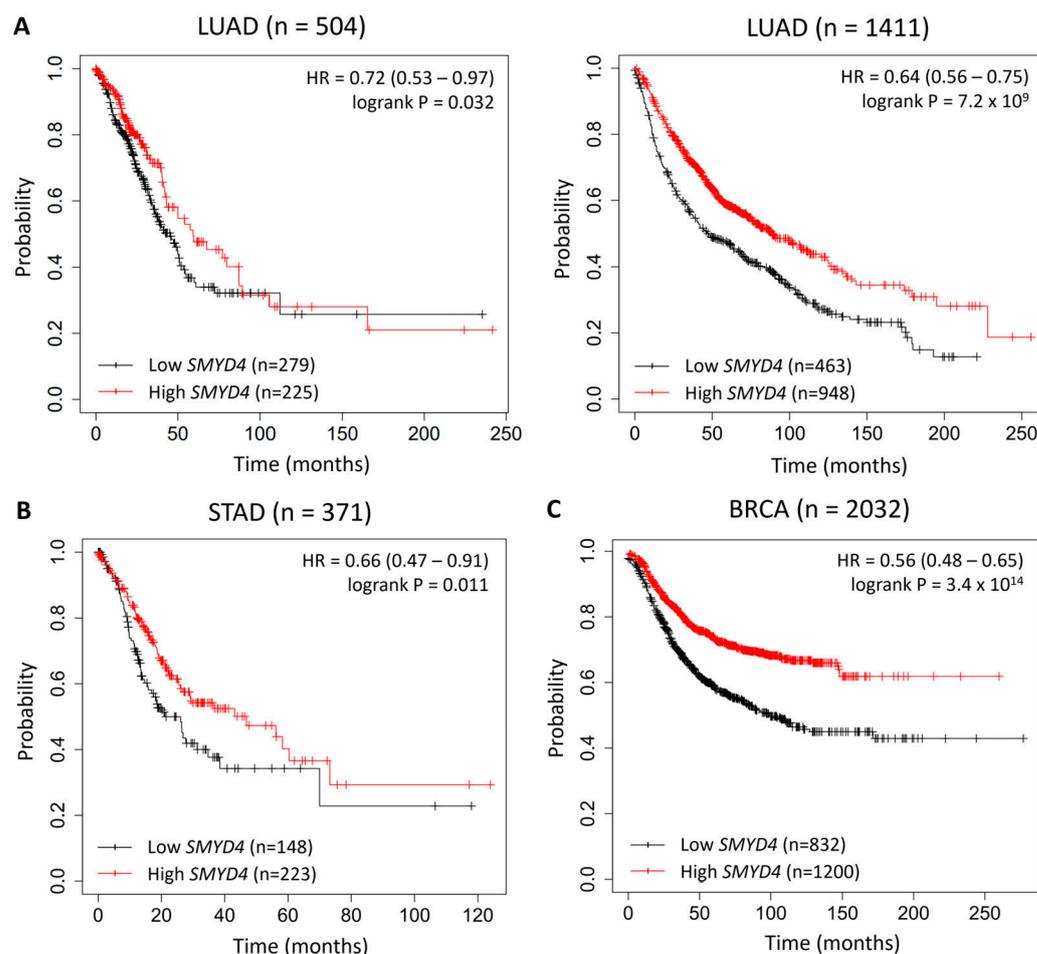
**Table 2.** Correlation analyses of *SMYD4* mRNA and CNA.

Tumors	Pearson	<i>p</i> Value	Spearman	<i>p</i> Value	Patients (n)
CRC	0.568	<0.0001	0.337	<0.0001	255
STAD	0.593	<0.0001	0.465	<0.0001	190
PRAD	0.468	<0.0001	0.296	0.0015	98
LUSC	0.615	<0.0001	0.500	<0.0001	362
LUAD	0.488	<0.0001	0.354	0.0003	89
BRCA	0.532	<0.0001	0.472	<0.0001	695
UCEC	0.483	<0.0001	0.538	<0.0001	74

### 2.7. Overall Patient Survival Based on *SMYD4* Expression

To assess whether there is a correlation between the overall survival (OS) of patients and the expression level of *SMYD4*, we analyzed its mRNA expression using both RNA-Seq data and microarray data (Affymetrix ID 229175\_at) available from the Kaplan–Meier plotter database. For each cancer type evaluated, we dichotomized patients based on *SMYD4* expression into two groups with high or low expression as previously described [12]. We found that low expression of *SMYD4* was significantly linked to a reduced OS in LUAD patients. In a cohort of 504 LUAD patients, those with lower expression of *SMYD4* had a median overall survival of 45.2 months compared with an overall survival of 59.2 months of patients with higher expression (HR = 0.72, 95% CI = 0.53–0.97;  $p = 0.032$ ) (Figure 6A). This difference became even more pronounced in a separate cohort of 1411 LUAD patients, where *SMYD4* expression was analyzed using microarrays. Patients exhibiting lower *SMYD4* expression had a median overall survival of 47 months. In contrast, patients with higher *SMYD4* expression demonstrated a median overall survival of 88.7 months (HR = 0.64, 95% CI = 0.56–0.75;  $p = 7.2 \times 10^{-9}$ ) (Figure 6A). Similarly, in a cohort of 371 STAD patients, those with higher expression had a median overall survival of 46.9 months, whereas those with lower *SMYD4* expression had a median overall survival of 21.1 months (HR = 0.66, 95% CI = 0.47–0.91;  $p = 0.011$ ) (Figure 6B).

When we analyzed available cohorts of BRCA patients, although *SMYD4* expression did not significantly affected OS, it was linked with a worse relapse-free survival (RFS). Patients who had lower expression of *SMYD4* demonstrated a median RFS of 28 months compared with the group that had higher expression (median RFS of 59 months, HR = 0.56 95% CI = 0.48–0.65;  $p = 3.4 \times 10^{-14}$ ) (Figure 6C).



**Figure 6.** Overall survival (OS) or relapse-free survival (RFS) analysis based on *SMYD4* expression levels (Kaplan-Meier plotter). **(A)** Two distinct cohorts of lung adenocarcinoma patients (LUAD) were analyzed. The left panel shows the OS in a cohort of 371 patients based on RNA-seq data. The right panel shows the OS of a cohort of 1411 patients based on microarray data. **(B)** OS in a cohort of 371 stomach adenocarcinoma (STAD) patients. **(C)** RFS in a cohort of 2032 breast cancer patients (n = 2031) was analyzed based on microarray data. Differences in OS or RFS were analyzed with the log-rank test.

### 3. Discussion

The availability of vast amounts of cancer genomic data in public database repositories has made in silico analysis an indispensable tool for exploring cancer-related vulnerabilities. Analyzing the most predominant genetic alterations in *SMYD* genes has revealed important distinctions among each member, their relation with cancer progression, and their usefulness as prognostic tools.

In the context of mutations, the missense variant emerges as the most frequently observed mutation across *SMYD* genes. The conserved SET domain, in particular, harbors a total of 107 missense mutations within the *SMYD* gene family. However, none of these mutations demonstrate a clear predominance, making it challenging to determine their potential impact on the functional dynamics of the *SMYD* family.

*SMYD4* emerged as a point of interest in our analysis, as all the evaluated tumors displayed a high frequency of heterozygous loss. This observation is consistent with the mRNA findings across all analyzed tumor types, which all showed a trend toward downregulation. This trend may be due to the fact that *SMYD4* is located on 17p13.3, a region known to undergo heterozygous loss in various solid tumors and leukemias [4,13–16]. Although not a rule, copy number alterations can affect gene expression [17]. The correlation analysis

between CNAs and mRNA expression in all evaluated tumors showed that both alterations in *SMYD4* are correlated.

Intriguingly, Xiao et al. proposed that *SMYD4* is accountable for di- and tri-methylation at H3K4 in zebrafish [18]. H3K4me3 is an epigenetic marker recognized for preserving the activity of tumor suppressor genes in normal cells [19]. Thus, if the function of *SMYD4* observed in zebrafish is conserved and mirrors its role in humans, *SMYD4* could potentially activate tumor suppressors through H3K4me3. Therefore, a loss in *SMYD4* function might lead to a corresponding loss of function in other tumor suppressors.

To date, very few studies have examined the role of *SMYD4* in cancer. Our analysis confirmed that *SMYD4* was downregulated in the majority of BRCA patients examined. Although there was no association with OS, downregulation of *SMYD4* significantly affected RFS. Importantly, BRCA patients also presented a significant correlation between CNAs and mRNA expression. This observation supports the findings that propose a tumor suppressor role for *SMYD4* in the development of breast cancer, at least partially, by inhibiting platelet-derived growth factor receptor  $\alpha$  polypeptide (Pdgfr- $\alpha$ ) [4]. Its downregulation also helps in the process of transforming normal mammary cells into tumor cells [5]. In addition, a study by Zhang et al. also found that *SMYD4* was downregulated in PRAD tumor tissues, which is also consistent with the findings of our study [20]. We also demonstrated that *SMYD4* consistently displayed a pattern of downregulation across a variety of tumor types, implying that its role as a tumor suppressor might be a universal characteristic in diverse oncological contexts.

Importantly, a comparative analysis of *SMYD4* expression between samples from normal tissues and samples from tumors at different stages revealed that *SMYD4* expression is significantly decreased in cancer samples, regardless of the tumor stage. Despite no association with tumor stage, *SMYD4* expression remains significantly reduced during cancer progression. This was observed in almost all cancer types analyzed, including colon and rectal cancers, breast cancer, uterine corpus endometrial carcinoma, and lung squamous cell carcinoma. For patients with lung adenocarcinoma, a decrease in *SMYD4* expression also resulted in a lower overall survival rate. This suggests that *SMYD4* expression could be explored as a valuable biomarker for predicting a more severe prognosis in this cancer type.

Interestingly, in the examined data from stomach adenocarcinoma, specifically those from stage 2, 3, and 4 tumors, *SMYD4* exhibited higher expression levels compared to normal samples in most patients. However, Kaplan–Meier analysis of 371 patients revealed that lower expression of *SMYD4* was associated with poorer overall survival, indicating a more unfavorable outcome among those patients with reduced *SMYD4* expression. One possible explanation for this intriguing alteration in *SMYD4* expression in STAD may be that higher *SMYD4* potentiates the expression of the transcription factor Nanog in cancer stem cells [7]. Previous studies have demonstrated a correlation between elevated Nanog expression in gastric tumors and increased tumor aggressiveness, as it enhances cell proliferation, migration and invasion [21,22]. A recent study also reports that *SMYD4* is upregulated in hepatocellular carcinoma, forming a positive feedback loop with the arginine methyltransferase PRMT5 [23]. As the role of *SMYD4* in cancer becomes more evident, further research is needed to better comprehend its mechanism of action in these tumor types.

The correlation analysis revealed noteworthy findings regarding changes in the expression pattern between *SMYD4* and other members of the *SMYD* family. In general, *SMYD4* negatively weak correlates with other *SMYD* genes. For instance, *SMYD3* is known to be upregulated in colorectal cancers [24–27]. In patients with high expression of *SMYD3*, *SMYD4* is consistently downregulated, exhibiting a clear negative correlation. Interestingly, a similar correlation was found in CRC between *SMYD4* and *SMYD5*, which is a much less studied gene. This negative correlation was also observed in PRAD, where patients with a high expression of *SMYD2*, *SMYD3* [26,28], or *SMYD5* showed lower expression of *SMYD4*. This opposing expression pattern was also observed in LUSC and BRCA, where *SMYD4* is downregulated while *SMYD5* is upregulated. In BRCA, there was a small negative

correlation between *SMYD4* and *SMYD3*, which is expected since *SMYD3* upregulation is already known to be associated with breast cancer proliferation [26,27,29,30]. Despite sharing similar conserved domains, the contrasting expression patterns of these genes in cancer may suggest divergent roles.

In conclusion, our study highlights *SMYD4* as a tumor suppressor gene across various solid tumors. Our comprehensive analysis of genetic alterations within the *SMYD4* gene revealed consistent heterozygous loss and downregulation across all tumors evaluated. Furthermore, *SMYD4* exhibited frequent negative correlations with other members of the *SMYD* family, suggesting distinct functional roles. Importantly, the consistent downregulation of *SMYD4* expression and its association with poor overall survival underscores its likely role in tumorigenesis and highlights its potential as a valuable biomarker. Further research into the mechanisms underlying *SMYD4*'s tumor-suppressive functions is warranted to fully elucidate its clinical potential.

## 4. Materials and Methods

### 4.1. Gene Database Analysis

Mutation data were obtained from the cBio Cancer Genomics Portal ([www.cbioportal.org](http://www.cbioportal.org)) and the Catalogue Of Somatic Mutations In *Cancer* (COSMIC) "[www.cancer.sanger.ac.uk/cosmic](http://www.cancer.sanger.ac.uk/cosmic)" (accessed on 31 October 2022)". Data on copy number alterations and mRNA expression were obtained from the cBio Cancer Genomics Portal "[www.cbioportal.org](http://www.cbioportal.org)" (accessed on 31 July 2022)" and the Cancer Genome Atlas (TCGA) database "[www.cancergenome.nih.gov](http://www.cancergenome.nih.gov)" (accessed on 31 June 2022). The frequency of copy number alterations (CNAs) was generated using the algorithm GISTIC (Genomic Identification of Significant Targets in Cancer), and the copy number of each gene in each tumor evaluated was determined, where  $-2$  indicates a homozygous loss,  $-1$  is heterozygous loss,  $1$  is low-level gain, and  $2$  indicates amplification. For mRNA expression analysis, data from patient samples relative to diploid samples were used. The value is presented as a standard deviation of the mean expression (Z-score) [20,31]. The UALCAN database "<http://ualcan.path.uab.edu>" (accessed on 31 August 2023)", which contains data obtained from TCGA [32], was used to evaluate the relative expression of *SMYD4* between normal and tumor stage samples in order to identify whether *SMYD4* could serve as a potential biomarker in the tumors evaluated in this study. Graphical representations were generated using GraphPad Prism 8 and Rstudio 2023.06.0 software.

### 4.2. Heatmap

The z-score data related to the mRNA expression of *SMYDs* in the evaluated cancer types were visualized using heatmap plots generated through the Flaski data analysis and visualization tool "<https://flaski.age.mpg.de>" (accessed on 30 September 2023)". The data were scaled numerically within a range of  $\pm 2$ . The Ward linkage method with Euclidean distance was employed for clustering. The data were sorted into row clusters.

### 4.3. Survival Curve

The Kaplan–Meier plotter database "<http://kmpplot.com/analysis/>" (accessed on 30 April 2023)" was used to explore the prognostic significance of alterations in the five *SMYD* genes. Survival analyses were conducted to determine the correlation between *SMYD* gene alterations and tumor prognosis. For mRNA expression analysis, the mRNA pan cancer option was selected, and patients were divided based on "high expression" and "low expression" levels as previously described [12]. OS for LUAD and PFS for BRCA were also assessed using data from microarray assays.  $p$  value  $< 0.05$  was considered statistically significant.

#### 4.4. Statistics

Statistical analyses were performed using GraphPad Prism 8. The mRNA correlations between *SMYD* family genes were analyzed using Pearson correlation test, with  $p < 0.05$  serving as the definition of a statistically significant difference.

**Author Contributions:** Conceptualization, B.L.O.S. and F.P.-S.; methodology, B.L.O.S. and M.B.d.L.; formal analysis, F.P.-S.; investigation, B.L.O.S.; resources, F.P.-S.; data curation, B.L.O.S., M.B.d.L. and F.P.-S.; writing—original draft preparation, B.L.O.S.; writing—review and editing, B.L.O.S., M.B.d.L., A.C.M.G. and F.P.-S.; visualization, B.L.O.S.; supervision, A.C.M.G. and F.P.-S.; project administration, F.P.-S.; funding acquisition, F.P.-S. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research was supported by Universidade de Brasília, Fundação de Amparo à Pesquisa do Distrito Federal (FAPDF), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). FP-S has received grants from FAPDF (grant numbers 00193-00001029/2021-95 and 00193-00002146/2023-38) and CNPq (grant numbers 440734/2022-3 and 406890/2022-6).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is contained within the article.

**Acknowledgments:** We would like to thank all the members of the Laboratory of Molecular Pathology of Cancer, Faculty of Health Sciences, University of Brasilia.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Albert, M.; Helin, K. Histone methyltransferases in cancer. *Semin. Cell Dev. Biol.* **2010**, *21*, 209–220. [[CrossRef](#)]
- Leinhart, K.; Brown, M. SET/MYND Lysine Methyltransferases Regulate Gene Transcription and Protein Activity. *Genes* **2011**, *2*, 210–218. [[CrossRef](#)]
- Liu, D.; Wang, X.; Shi, E.; Wang, L.; Nie, M.; Li, L.; Jiang, Q.; Kong, P.; Shi, S.; Wang, C.; et al. Comprehensive Analysis of the Value of SMYD Family Members in the Prognosis and Immune Infiltration of Malignant Digestive System Tumors. *Front. Genet.* **2021**, *12*, 699910. [[CrossRef](#)]
- Hu, L.; Zhu, Y.T.; Qi, C.; Zhu, Y.J. Identification of Smyd4 as a potential tumor suppressor gene involved in breast cancer development. *Cancer Res.* **2009**, *69*, 4067–4072. [[CrossRef](#)]
- Han, S.; Zou, H.; Lee, J.W.; Han, J.; Kim, H.C.; Cheol, J.J.; Kim, L.S.; Kim, H. miR-1307-3p Stimulates Breast Cancer Development and Progression by Targeting SMYD4. *J. Cancer* **2019**, *10*, 441–448. [[CrossRef](#)]
- Aziz, N.; Hong, Y.H.; Kim, H.G.; Kim, J.H.; Cho, J.Y. Tumor-suppressive functions of protein lysine methyltransferases. *Exp. Mol. Med.* **2023**, *55*, 2475–2497. [[CrossRef](#)]
- Liu, S.; Cheng, K.; Zhang, H.; Kong, R.; Wang, S.; Mao, C.; Liu, S. Methylation Status of the Nanog Promoter Determines the Switch between Cancer Cells and Cancer Stem Cells. *Adv. Sci.* **2020**, *7*, 1903035. [[CrossRef](#)]
- Sung, H.; Ferlay, J.; Siegel, R.L.; 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 Cancer J Clin.* **2021**, *71*, 209–249. [[CrossRef](#)]
- Kastenhuber, E.R.; Lowe, S.W. Putting p53 in Context. *Cell* **2017**, *170*, 1062–1078. [[CrossRef](#)]
- Albertson, D.G.; Collins, C.; McCormick, F.; Gray, J.W. Chromosome aberrations in solid tumors. *Nat. Genet.* **2003**, *34*, 369–376. [[CrossRef](#)]
- Mermel, C.H.; Schumacher, S.E.; Hill, B.; Meyerson, M.L.; Beroukhim, R.; Getz, G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **2011**, *12*, R41. [[CrossRef](#)]
- Lánczky, A.; Györfy, B. Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. *J. Med. Internet Res.* **2021**, *23*, e27633. [[CrossRef](#)]
- Jenal, M.; Britschgi, C.; Fey, M.F.; Tschan, M.P. Inactivation of the hypermethylated in cancer 1 tumour suppressor—not just a question of promoter hypermethylation? *Swiss Med. Wkly.* **2010**, *140*, w13106. [[CrossRef](#)]
- Wu, G.J.; Shan, X.N.; Li, M.F.; Shi, S.L.; Zheng, Q.P.; Yu, L.; Zhao, S.Y. Preliminary study on the loss of heterozygosity at 17p13 in gastric and colorectal cancers. *World J. Gastroenterol.* **1997**, *3*, 160–162. [[CrossRef](#)]
- Tsuchiya, E.; Tanigami, A.; Ishikawa, Y.; Nishida, K.; Hayashi, M.; Tokuchi, Y.; Hashimoto, T.; Okumura, S.; Tsuchiya, S.; Nakagawa, K. Three New Regions on Chromosome 17p13.3 Distal to p53 with Possible Tumor Suppressor Gene Involvement in Lung Cancer. *Jpn. J. Cancer Res.* **2000**, *91*, 589–596. [[CrossRef](#)]
- Park, S.Y.; Kang, Y.S.; Kim, B.G.; Lee, S.H.; Lee, E.D.; Lee, K.H.; Park, K.B.; Lee, J.H. Loss of heterozygosity on the short arm of chromosome 17 in uterine cervical carcinomas. *Cancer Genet. Cytogenet.* **1995**, *79*, 74–78. [[CrossRef](#)]

17. Bhattacharya, A.; Bense, R.D.; Urzúa-Traslaviña, C.G.; de Vries, E.G.; van Vugt, M.A.; Fehrmann, R.S. Transcriptional effects of copy number alterations in a large set of human cancers. *Nat. Commun.* **2020**, *11*, 715. [CrossRef]
18. Xiao, D.; Wang, H.; Hao, L.; Guo, X.; Ma, X.; Qian, Y.; Chen, H.; Ma, J.; Zhang, J.; Sheng, W.; et al. The roles of SMYD4 in epigenetic regulation of cardiac development in zebrafish. *PLoS Genet.* **2018**, *14*, e1007578. [CrossRef]
19. Chen, K.; Chen, Z.; Wu, D.; Zhang, L.; Lin, X.; Su, J.; Rodriguez, B.; Xi, Y.; Xia, Z.; Chen, X.; et al. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat. Genet.* **2015**, *47*, 1149–1157. [CrossRef]
20. Zhang, Y.; Yan, L.; Yao, W.; Chen, K.; Xu, H.; Ye, Z. Integrated Analysis of Genetic Abnormalities of the Histone Lysine Methyltransferases in Prostate Cancer. *Med. Sci. Monit.* **2019**, *25*, 193–239. [CrossRef]
21. Lin, T.; Ding, Y.Q.; Li, J.M. Overexpression of Nanog protein is associated with poor prognosis in gastric adenocarcinoma. *Med. Oncol.* **2012**, *29*, 878–885. [CrossRef]
22. Ma, X.; Wang, B.; Wang, X.; Luo, Y.; Fan, W. NANOGP8 is the key regulator of stemness, EMT, Wnt pathway, chemoresistance, and other malignant phenotypes in gastric cancer cells. *PLoS ONE* **2018**, *13*, e0192436. [CrossRef]
23. Zhou, Z.; Chen, Z.; Zhou, Q.; Meng, S.; Shi, J.; Mui, S.; Jiang, H.; Lin, J.; He, G.; Li, W.; et al. SMYD4 monomethylates PRMT5 and forms a positive feedback loop to promote hepatocellular carcinoma progression. *Cancer Sci.* **2024**, *115*, 1587–1601. [CrossRef]
24. Hamamoto, R.; Furukawa, Y.; Morita, M.; Iimura, Y.; Silva, F.P.; Li, M.; Yagyu, R.; Nakamura, Y. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat. Cell Biol.* **2004**, *6*, 731–740. [CrossRef]
25. Liu, N.; Sun, S.; Yang, X. Prognostic significance of stromal SMYD3 expression in colorectal cancer of TNM stage I-III. *Int J Clin Exp Pathol.* **2017**, *10*, 8901–8907.
26. Bernard, B.J.; Nigam, N.; Burkitt, K.; Saloura, V. SMYD3: A regulator of epigenetic and signaling pathways in cancer. *Clin. Epigenetics* **2021**, *13*, 45. [CrossRef]
27. Bottino, C.; Peserico, A.; Simone, C.; Caretti, G. SMYD3: An Oncogenic Driver Targeting Epigenetic Regulation and Signaling Pathways. *Cancers* **2020**, *12*, 142. [CrossRef]
28. Ikram, S.; Rege, A.; Negesse, M.Y.; Casanova, A.G.; Reynoird, N.; Green, E.M. The SMYD3-MAP3K2 Signaling Axis Promotes Tumor Aggressiveness and Metastasis in Prostate Cancer. *Sci. Adv.* **2023**, *9*, eadi5921. Available online: <https://www.science.org/doi/10.1126/sciadv.adi5921> (accessed on 28 April 2024). [CrossRef]
29. Hamamoto, R.; Silva, F.P.; Tsuge, M.; Nishidate, T.; Katagiri, T.; Nakamura, Y.; Furukawa, Y. Enhanced SMYD3 expression is essential for the growth of breast cancer cells. *Cancer Sci.* **2006**, *97*, 113–118. [CrossRef]
30. Fenizia, C.; Bottino, C.; Corbetta, S.; Fittipaldi, R.; Floris, P.; Gaudenzi, G.; Carra, S.; Cotelli, F.; Vitale, G.; Caretti, G. SMYD3 promotes the epithelial-mesenchymal transition in breast cancer. *Nucleic Acids Res.* **2019**, *47*, 1278–1293. [CrossRef]
31. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404. [CrossRef] [PubMed]
32. Chandrashekar, D.S.; Karthikeyan, S.K.; Korla, P.K.; Patel, H.; Shovon, A.R.; Athar, M.; Netto, G.J.; Qin, Z.S.; Kumar, S.; Manne, U.; et al. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia* **2022**, *25*, 18–27. [CrossRef] [PubMed]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.