

LUÍS AUGUSTO MUNIZ TELLES

ANÁLISE DE EXPRESSÃO DA METILTRANSFERASE SETD4 EM LEUCEMIA
LINFOIDE AGUDA E SUA RELAÇÃO COM A LEUCEMOGÊNESE

BRASÍLIA, 2016

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

LUÍS AUGUSTO MUNIZ TELLES

ANÁLISE DE EXPRESSÃO DA METILTRANSFERASE *SETD4* EM LEUCEMIA
LINFOIDE AGUDA E SUA RELAÇÃO COM A LEUCEMOGÊNESE

Dissertação apresentada como requisito parcial
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: Fábio Pittella Silva

Co-orientador: Luís Henrique Toshihiro Sakamoto

BRASÍLIA

2016

LUÍS AUGUSTO MUNIZ TELLES

PERFIL DE EXPRESSÃO DA METILTRANSFERASE *SETD4* E SUAS
INTERAÇÕES COM OUTRAS PROTEÍNAS

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

Aprovado em 9 de agosto de 2016

BANCA EXAMINADORA

Fábio Pittella Silva – Presidente

Universidade de Brasília, UnB

Andrea Barreto Motoyama

Universidade de Brasília, UnB

Felipe Saldanha de Araújo

Universidade de Brasília, UnB

Diêgo Madureira de Oliveira - Suplente

Universidade de Brasília, UnB

*Dedico este trabalho ao Único digno... o
Senhor, o meu Pai, o meu Deus.*

AGRADECIMENTOS

Agradeço ao meu Senhor, meu querido Deus. Ele é quem me sustentou todo o tempo em que cursei esta pós-graduação e quem esteve comigo nos momentos mais alegres e nos mais chorosos também. Ele é quem me ensinou e repreendeu nos momentos em que necessitei e quem me deu forças e encorajou quando eu tive medo. Agradeço por Ele ter posto pessoas propositalmente na minha vida. Dentre elas, está o meu orientador, Prof. Fábio. Agradeço por ter me aceitado no laboratório quando era ainda um aluno de PIBIC e por ter me dado a chance de aprender muitas coisas novas, dentre as quais se destaca a análise dos dados e a biologia computacional, que acabaram por me conquistar. Agradeço aos queridos colegas de laboratório que estiveram comigo nas conversas, nas ajudas com soluções, experimentos, desabafos, nas situações extremas eventuais que passamos nestes últimos dois anos e tantos outros momentos especiais também. Dentre eles, cito a Dora, que me ajudou bastante em vários momentos dessa caminhada; a Vivian, que chegou há pouco, mas que já se tornou muito mais que uma colega, mas uma irmã em Cristo e amiga; a Prof. Andrea, que também me ajudou muitas vezes e com muita paciência; os professores Felipe e Rodrigo, que foram bons companheiros de nitrogênio muitas vezes; o Rubens e a Martha, que foram pacientes e muito solícitos quando eu ainda era um aluno de PIBIC; o Lúcio, que também dividiu muitas conversas, almoços e planos; e o Orlene, que dividiu comigo experiências de vida e filosofias. Agradeço ao Luís Sakamoto, pois sem ele, de maneira alguma este projeto e muito menos esta dissertação existiriam. Agradeço ao Pedro Alexandre, querido irmão em Cristo, que me ajudou demais e também me estimulou com a linguagem R quando eu ainda era iniciante e mesmo agora no final. Por fim, agradeço ao núcleo sólido da minha vida depois de Deus: minha igreja e minha família. A primeira, família pela fé, e a segunda, família pelo sangue. Sem o apoio constante de meus irmãos em Cristo, sem o sustento, a paciência, a cooperação dos meus pais, sem minha irmã querida que está sempre comigo... sem eles, eu também jamais teria chegado até aqui. Obrigado queridos irmãos que estiveram comigo em oração. Obrigado, meu pai e minha mãe. Obrigado, Carol. Obrigado, João Victor. E agradeço também finalmente a minha linda amiga, Simone, que chegou mais proximamente há pouco tempo, mas como me tem ajudado até aqui!

*“Cristo, em quem estão escondidos todos os
tesouros da sabedoria e da ciência”*

Colossenses 2:2,3

RESUMO

A leucemia linfoblástica aguda (LLA) é a neoplasia com maior prevalência na infância. Melhorias muito expressivas no prognóstico da doença têm sido apresentadas nas últimas décadas, alcançando-se taxas de cura em torno de 90% em alguns casos. No entanto, a doença ainda apresenta mau prognóstico em pacientes adultos ou menores de um ano. A compreensão das neoplasias tem sido reinterpretada à luz do advento da epigenética como uma nova área do conhecimento. Constituindo um padrão de herança informacional que não depende de alterações na cadeia primária de DNA, a epigenética tem aberto o caminho para novas possibilidades de tratamento do câncer. As metiltransferases de lisina, em particular, têm estado em foco devido aos crescentes relatos de envolvimento na regulação da expressão de oncogenes ou genes supressores de tumor, bem como devido ao desenvolvimento de fármacos cujos alvos são estas mesmas enzimas. A família de metiltransferases SETD possui alguns membros que já foram descritos como tendo relacionamento com o câncer. No entanto, *SETD4* ainda carece de mais profunda caracterização. Nesse sentido, este trabalho teve como objetivo descrever o perfil de transcrição de *SETD4* em amostras clínicas de aspirado de medula óssea por meio de qPCR e verificar a sua relação com a leucemogênese por meio de análise de correlação com fatores clínicos e com a transcrição de outros genes. Também foi realizada uma análise exploratória de dados de expressão e transcrição disponíveis em bancos de bioinformática on-line. Foi verificado que *SETD4* encontra-se com transcrição aumentada no grupo de amostras clínicas neoplásicas em relação às amostras não-neoplásicas de nossa coorte, exibindo aumento transcricional de cerca de 5.4 vezes. Durante o tratamento quimioterápico, os pacientes exibiram redução da transcrição da metiltransferase, revelando correlação com a queda no número de linfoblastos na medula óssea no 29º dia de indução quimioterápica. *SETD4* teve sua transcrição fortemente correlacionada com a de duas outras metiltransferases: *SETMAR* e *SMYD2*. Foi detectada transcrição aumentada de *SETD4* em dados obtidos por microarranjo e RNA-seq de outras neoplasias em plataformas disponíveis de dados públicos. Além disso, amplificações e deleções de número de cópias do gene parecem influenciar negativamente a sobrevida dos pacientes de leucemia mieloide aguda, e câncer de estômago, enquanto o nível aumentado de transcrição parece ter o mesmo efeito em LLA.

Também verificamos que SETD4 apresenta mutações no domínio SET e Rubs-sub-bind que têm o potencial de impactar sua função enzimática. Juntos, estes resultados apontam para essa metiltransferase como uma enzima que pode ter um papel importante na leucemogênese e demais processos de carcinogênese, constituindo um atraente alvo terapêutico e farmacológico.

Palavras-chave: SETD4; neoplasias; leucemia-linfoma linfoblástico de células precursoras; epigenética; metiltransferases

ABSTRACT

Acute lymphoblastic leukemia (ALL) is the most prevalent cancer in childhood. Very significant improvements in the prognosis of the disease have been made in recent decades, with cure rates being achieved around 90% in some cases. However, the disease still has poor prognosis in adult patients or children with less than one year. The understanding of cancer has been reinterpreted in the light of the advent of epigenetics as a new area of knowledge. Constituting a pattern of informational inheritance that does not depend on changes on primary chain of DNA, epigenetics has opened the way to new possibilities of cancer treatment. Lysine methyltransferases, in particular, have been in focus due to the increasing reports of its involvement in the regulation of oncogene or tumor suppressor gene expression, as well as to the development of drugs targeting those same proteins. The SETD methyltransferase family has some members who have been described as being related to cancer. Despite all these efforts, *SETD4* still lacks deeper characterization. Therefore, this study aimed to describe *SETD4* transcription profile in clinical samples of bone marrow aspirate by qPCR and to assess its relationship with leukemogenesis by means of correlation analysis with clinical factors and transcription of other genes during chemotherapy. An exploratory analysis of data available at online bioinformatics platforms was also performed. It was found that *SETD4* transcription is upregulated in the neoplastic group of clinical samples relative to non-neoplastic samples, exhibiting an increase of approximately 5.4 times. During chemotherapy, patients exhibited reduced methyltransferase transcription, showing correlation with the decrease in the number of lymphoblasts in the bone marrow. *SETD4* was highly correlated with the transcription of two other methyltransferases: *SETMAR* and *SMYD2*. Upregulated transcription of *SETD4* was detected in data obtained by microarray and RNA-seq of other neoplasias in public data platforms. Moreover, gene copy number amplifications and deletions seems to have a negative influence on survival of acute myeloid leukemia and gastric cancer patients while upregulation of transcription in ALL seems to have the same effect. We also verified that *SETD4* presents mutations in SET and Rubs-subsub-bind domains that may have an impact on its enzymatic function. Together, these results point to this methyltransferase as an enzyme that may play a role in leukemogenesis and other

carcinogenesis processes, constituting an attractive therapeutic and pharmacological target.

Keywords: *SETD4*; neoplasms; precursor cell lymphoblastic leukemia-lymphoma; epigenetics; methyltransferases

LISTA DE FIGURAS

Figura 1. Estrutura do gene, transcrito e proteína SETD4.

Figura 2. Expressão de mRNA de *SETD4* obtida por RT-qPCR em amostras leucêmicas comparadas a amostras de medula óssea não-neoplásicas.

Figura 3. Expressão de mRNA de *SETD4* obtida pela plataforma de dados de microarranjo Oncomine™.

Figura 4. Níveis de transcrição de *SETD4* obtidos por microarranjo em amostras de LLA e de medula óssea normal.

Figura 5. Transcrição de *SETD4* em amostras de pacientes de LLA comparada à diminuição do fardo leucêmico durante a quimioterapia.

Figura 6. Expressão de mRNA de SETMAR em amostras leucêmicas comparadas a amostras de medula óssea não-neoplásicas.

Figura 7. Correlação entre a expressão de *SETD4*, *SMYD2* e *SETMAR* durante o tratamento quimioterápico.

Figura 8. Análise de sobrevida em pacientes de LLA estratificados por categoria de expressão de *SETD4*.

Figura 9. Curvas de sobrevida em pacientes de várias neoplasias estratificados por categoria de CNAs em *SETD4*.

Figura 10. Níveis de mRNA de *SETD4* em tecidos normais e em linhagens celulares de câncer dos respectivos tecidos.

Figura 11. Níveis de transcrição de *SETD4* em medula óssea comparada a linhagens de linfoma, de mieloma múltiplo e de leucemia.

Figura 12. Expressão de *SETD4* em amostras de várias neoplasias obtidos na *Cancer Cell Line Encyclopedia*.

Figura 13. Expressão de *SETD4* por RNA-seq em diversos tecidos não neoplásicos humanos.

Figura 14. Gráfico de pizza ilustrando a proporção das frequências de cada mutação de *SETD4*.

Figura 15. Diagrama de mutações em *SETD4* em amostras de vários tipos de câncer obtido pela plataforma cBioPortal.

LISTA DE TABELAS

Tabela 1. Características clínicas da coorte ao diagnóstico e quimioterapia pós-indutória.

Tabela 2. Fatores clínicos estratificados por categoria de expressão de *SETD4*.

Tabela 3. Informações sobre picos de amplificação/deleção abrangendo *SETD4*.

Tabela 4. Mutações de *SETD4* registradas na plataforma COSMIC.

Tabela 5. Predição de impacto de mutações *missense* na função da proteína *SETD4* com informações de impacto geradas pelo *Mutation Assessor* (cBioPortal).

LISTA DE ABREVIATURAS E SIGLAS

ABL1 – Homólogo 1 do oncogene viral leucemia murina de Abelson (do inglês, *Abelson murine leukemia viral oncogene homolog 1*)

Acetil-coA – Acetilcoenzima A

ACTB – Gene beta-actina

B-LLA – Leucemia linfoblástica aguda de células B

BCR – Gene *breakpoint cluster region*

CCLE – Plataforma de dados *Cancer Cell Line Encyclopedia*

CD – Cluster of differentiation

cDNA – DNA complementar

CEP-FEPECS – Comitê de Ética em Pesquisa da Fundação de Ensino e Pesquisa em Ciências Da Saúde

CGI – Ilha de dinucleotídeos CpG

CNA – Alteração de número de cópias (do inglês, *copy number alteration*)

CpG – Dinucleotídeo citosina fosfato guanina

CO₂ – Gás carbônico

CpG – Dinucleotídeo Citosina-Fosfato-Guanina

Cq – Ciclo de quantificação

CRLF2 – Fator semelhante a receptor de citocina 2 (do inglês, *cytokine receptor-like factor 2*)

CREB – Proteína ligante ao elemento responsivo a AMP cíclico 1 (do inglês, *cyclic AMP responsive elemento binding protein*)

CREBBP – Proteína ligante a CREB

DNA – Ácido desoxirribonucleico (do inglês, *deoxiribonucleic acid*)

DNMT – Metiltransferase de DNA (do inglês, *DNA methyltransferase*)

DOT1L – Gene *Disruptor of telomere silencing 1-like*

EBF1 – Fator precoce de células B 1 (do inglês, *early B-cell factor 1*)

ELSPB1 – Proteína ligante de esperma epididimal (do inglês, *epididymal sperm binding protein 1*), também conhecida como E12

EST – *Expressed Sequence Tag*

ETV6 – Variante do gene ETS 6 (do inglês, *ETS variant 6*), também conhecido como TEL (do inglês *translocation-ETS-leukemia*)

EPOR – Receptor de eritropoietina (do inglês, *erythropoietin receptor*)

ER – Receptor de estrógeno

EUA – Estados Unidos da América

FISH - Hibridização fluorescente *in situ*

FPKM – Fragments per kilobase of exon per million fragments mapped

GBTLI – Grupo Brasileiro de Tratamento de Leucemia da Infância

GNAT – Família de acetiltransferases de lisina nominada após seu membro Gcn5 (scKAT2)

GTE_x – Plataforma de dados Genotype-Tissue Expression

H1 – Histona H1

H2A – Histona H2A

H2B – Histona H2B

H3 – Histona H3

H4 – Histona H4

HAT – Acetiltransferase de histonas (do inglês, *histone acetyltransferase*)

HCC – Carcinoma hepatocelular (do inglês, *hepatocellular carcinoma*)

HDAC – Deacetilase de histonas (do inglês, *histone deacetylase*)

HepG2 – Linhagem celular de carcinoma hepatocelular

HoxA – *Cluster* de genes *homeobox A*

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

ID - Identificador

IKZF1 – Proteína dedo de zinco 1 da família IKAROS (do inglês, *IKAROS family zinc finger 1*)

iRNA – RNA interferente

JAK – Janus quinase

KAT – Acetiltransferase de lisina (do inglês, *lysine acetyltransferase*)

KMT2A – Acetiltransferase de lisina 2^a

LLA – Leucemia linfoblástica aguda, pode ser de células B (B-LLA) ou de células T (T-LLA)

LLC – Leucemia linfoblástica crônica

LMA – Leucemia mieloide aguda

LMC – Leucemia mieloide crônica

lncRNA – RNA longo não codante

LPMC – Laboratório de Patologia Molecular do Câncer

LSMT – Domínio metiltransferase da subunidade grande da proteína Rubisco (do inglês, *large subunit methyltransferase*)

MDA-MB 231 – Linhagem celular de carcinoma mamário

MI – Estado de *Michigan*, nos EUA

MILE – Estudo Inovações de Microarranjo em Leucemia (do inglês, *Microarray Innovations in Leukemia*)

miRNA – RNA pertencente à classe dos micro RNA's

MIT – Instituto de Tecnologia de Massachussets (do inglês, *Massachussets Institute of Technology*)

MLL – Gene *mixed-lineage leukemia*, também conhecido como KMT2A (do inglês, *lysine methyltransferase 2A*)

mRNA – RNA mensageiro

MYST – Família de acetiltransferases nominada após seus membros MOZ, Ybf2/Sas3, Sas2 e Tip60

NAD⁺ – Dinucleotídeo de nicotinamida e adenina (do inglês, *nicotinamide and adenine dinucleotide*)

NCBI – Centro Nacional para Informação em Biotecnologia (do inglês, *National Center for Biotechnology Information*)

NHEJ – Junção de finais não homóloga (do inglês *non-homologous end junction*)

NT5C2 – Nucleotidase 5' citosólica II (do inglês, 5'-nucleotidase, cytosolic II)

OAADPr – O-acetil-ADP-ribose

p53 – Proteína supressora tumoral p53

p300/CBP – Família de acetiltransferases

PAX5 – Proteína caixa pareada 5 (do inglês, *paired box 5*)

PBMC – Células mononucleares do sangue periférico (do inglês, *peripheral blood mononuclear cells*)

PBS – Tampão fosfato salino

PBX1 – Gene homeobox PBX1

PDGFR – Receptor de fator de crescimento derivado de plaquetas (do inglês, *platelet derived growth factor receptor*)

piRNA – RNA de interação com *piwi*

PPARGC1A - Coativador 1 alfa de PPARG (receptor ativado por proliferadores de peroxissoma gama, do inglês *peroxisome proliferator activated receptor gamma*)

PTM – Modificação pós-traducional (do inglês, *post-translational modification*)

RAS – Proteína vírus de sarcoma de ratos (do inglês, *rat sarcoma virus*)

RB1 – Gene retinoblastoma 1

RMA – Técnica de normalização média robusta multiarranjo (do inglês *multiarray robust average*)

RNA – Ácido ribonucleico

RNA-seq – Sequenciamento de RNA

RQ – Quantificação relativa (do inglês, *relative quantification*)

RT-qPCR – Reação em cadeia de polimerase de transcrição reversa quantitativa

Rubs-subst-bind – Domínio de ligação ao substrato da proteína Rubisco

RUNX1 – Gene fator de transcrição relacionado a *run 1*, também conhecido como leucemia mieloide aguda 1 (AML1; do inglês *acute myeloid leukemia 1*)

SAM – S-adenosil-metionina

SET – Família SET de metiltransferases

SETD3 - Gene *SET domain containing 3*

SETD4 – Gene *SET domain containing 4*

SETD6 – Gene *SET domain containing 6*

SETD7 – Gene *SET domain containing 7*

SETD8 – Gene *SET domain containing 8*

SETMAR – Gene *SET domain and mariner transposase fusion gene*

shRNA – RNA do tipo *short hairpin*

SISNEP – Sistema Nacional de Ética em Pesquisa

SMYD2 – Gene *SET and MYND domain containing 2*

snoRNA – RNA nucleolar pequeno

T-LLA – Leucemia Linfoblástica Aguda de Células T

TCF3 – Fator de transcrição 3 (do inglês, *transcription factor 3*), também conhecido como E47

TET – Proteína Ten-Eleven-Translocation

UCSC – Universidade da Califórnia Santa Cruz

UnB – Universidade de Brasília

WHSC1 - Gene *Wolf-Hirschhorn syndrome candidate 1*, antes conhecido como *nuclear receptor binding SET domain containg 2 (NSD2)* ou *multiple myeloma SET domain (MMSET)*

SUMÁRIO

1 INTRODUÇÃO	1
1.1 LEUCEMIA LINFOBLÁSTICA AGUDA.....	1
1.1.1 Subtipos de B-LLA e alterações citogenéticas específicas	2
1.1.1.1 Fusão <i>ETV6-RUNX1</i>	2
1.1.1.2 Fusão <i>BCR-ABL1</i> (Cromossomo Filadélfia).....	3
1.1.1.3 Fusão <i>TCF3-PBX1</i>	3
1.1.1.4 Rearranjo de <i>MLL</i>	3
1.1.1.5 Subtipo com Hiperdiploidia e subtipo com Hipodiploidia.....	4
1.1.1.6 Subtipos de alto risco.....	4
1.1.1.6.1 <i>LLA De Precursores Jovens De Células T</i>	4
1.1.1.6.2 <i>LLA Semelhante a BCR-ABL1⁺</i>	5
1.1.1.6.3 <i>LLA associada a Síndrome de Down</i>	5
1.2 EPIGENÉTICA E CÂNCER.....	6
1.2.1 Epigenética	6
1.2.1.1 Metilação de DNA.....	8
1.2.1.2 Alterações pós-traducionais em histonas.....	9
1.2.1.2.1 <i>Acetilação de Histonas</i>	9
1.2.1.2.2 <i>Metilação de Histonas</i>	10
1.2.1.3 Outros Mecanismos Epigenéticos.....	12
1.2.1.4 Alterações Epigenéticas no Câncer.....	12
1.2.1.5 Metiltransferase SETD4.....	13
2 OBJETIVOS	16
2.1 OBJETIVO GERAL.....	16
2.2 OBJETIVOS ESPECÍFICOS.....	16
3 MÉTODOS	17
3.1 AMOSTRAS CLÍNICAS DE LLA, ENSAIO DE RT-qPCR E ANÁLISE DE SOBREVIDA.....	17
3.2 ANÁLISE EXPLORATÓRIA DE DADOS DE REPOSITÓRIOS ON-LINE DE BIOINFORMÁTICA.....	18
4 RESULTADOS	20
4.1 EXPERIMENTOS COM AMOSTRAS CLÍNICAS.....	20
4.1.1 SETD4 Apresenta Aumento de Expressão em Amostras de LLA Infantil	20
4.1.2 A Expressão de SETD4 Diminui Durante a Quimioterapia, Bem Como o Número de Células Leucêmicas	26
4.1.3 Os Níveis de Expressão de SETD4, SETMAR e SMYD2 Estão Correlacionados Durante o Tratamento Quimioterápico	29
4.2 ANÁLISES EM BANCOS DE DADOS.....	32
4.2.1 Análise de Sobrevida com Dados do cBioPortal de Pacientes Estratificados por Alterações de Número de Cópias (CNA's)	33
4.2.2 Análise de Dados de transcrição de SETD4 no The Human Protein Atlas	35

4.2.3 Análise de Dados no Tumorscape.....	37
4.2.4 Análise de Dados de Expressão na <i>Cancer Cell Line Encyclopedia</i> (CCLE)	38
4.2.5 Análise de Dados de Expressão no GTEx Portal	39
4.2.6 Análise de Mutações na plataforma COSMIC e cBioPortal.....	40
5 DISCUSSÃO.....	43
6 CONCLUSÃO.....	47
REFERÊNCIAS.....	48
ANEXO I - TABELA A1 - NÚMERO AMOSTRAL DE CADA TECIDO ANALISADO NA PLATAFORMA GTEX	56
ANEXO II - TABELA A2 - VALORES DE P OBTIDOS PELA APLICAÇÃO DO TESTE DE LOG-RANK PARA TENDÊNCIAS ÀS CURVAS OBTIDAS PELO CBIOPORTAL	57

1 INTRODUÇÃO

1.1 LEUCEMIA LINFOBLÁSTICA AGUDA

Leucemia é o nome dado às neoplasias do tecido hematopoiético que acometem primariamente a medula óssea. De acordo com a linhagem celular de origem, a leucemia é classificada como mieloide ou linfoide, e de acordo com o curso da doença, a leucemia é classificada como sendo crônica ou aguda, de maneira que há quatro tipos principais de leucemia: a leucemia mielóide crônica (LMC), a leucemia mielóide aguda (LMA), a leucemia linfoblástica crônica (LLC) e a leucemia linfoblástica aguda (LLA) (1).

A LLA constitui uma neoplasia que emerge de células precursoras de linfócitos e é agrupada em dois grandes tipos: a doença de células B (B-LLA), que não inclui a leucemia/linfoma de Burkitt, um tumor de células maduras, e a de células T (T-LLA) (2–4). Com um pico de incidência na faixa etária entre 2 a 5 anos de idade (excetuando-se o subtipo com rearranjo de MLL), a LLA é tida como a neoplasia infantil mais comum, contabilizando cerca de um quarto de todos os casos de câncer infantil (3,5). Desta parcela, a B-LLA e as leucemias de linhagens mistas compreendem 85% e a T-LLA entre 10 a 15% dos casos (3).

No passado, pacientes de LLA eram vítimas de um prognóstico ruim, apresentando índices de sobrevivência tão pequenos quanto 10% na década de 1960 (6,7). Felizmente, este cenário mudou muito com o passar dos anos e a melhora nos cuidados de suporte que os seguiram. Desde a identificação de aberrações cromossomais por técnicas citogenéticas convencionais, o advento de novas tecnologias como o sequenciamento de nova geração (do genoma e do transcriptoma), microarranjos, a hibridização fluorescente *in situ* (FISH), detecção de alterações epigenéticas e de número de cópias de DNA, possibilitou a detecção de subtipos da neoplasia caracterizados por alterações cromossomais grosseiras ou por aquelas alterações genéticas de nível submicroscópico, as quais foram reconhecidas como importantes fatores de prognóstico (3–5). Assim, desenvolveu-se uma estratificação de risco e terapias ajustadas ao contexto genético do paciente.

Adicionalmente, houve o desenvolvimento de tratamentos quimioterápicos multiagentes e o reconhecimento do sistema nervoso central e dos testículos como “santuários leucêmicos” (sítios que apresentam maior risco de recaída), possibilitando o direcionamento de terapias a estes locais (8,9). Juntos, todos estes fatores tem contribuído para a melhora da sobrevida dos pacientes de LLA, obtendo-se índices de cerca de 90% de cura em registros mais atuais (6,7,10) e mudando-se o quadro de pacientes com fatores antes considerados de risco, tais como o sexo masculino e a raça negra (4,11).

Atualmente, muitos subtipos de B-LLA estão bem estabelecidos e relacionados a estratificação de risco, enquanto, para T-LLA, são definidos quatro grandes subtipos. Para a última, aberrações citogenéticas não tem relação confirmada com diferentes prognósticos.

1.1.1 Subtipos de LLA e alterações citogenéticas específicas

Aberrações citogenéticas são uma característica comum em LLA e têm sido conhecidas há décadas. As translocações estão presentes em grande parte dos casos e é provável que representem aqueles eventos iniciadores pré-leucêmicos, havendo a necessidade de um evento adicional para o aparecimento da leucemia propriamente dita. Sabe-se, devido a estudos com gêmeos monozigóticos e estudos com teste de Guthrie, que estas alterações têm uma origem pré-natal na maioria dos casos (12,13). Junto com as alterações de número de cromossomos, elas constituem fatores que determinam o prognóstico da doença e possibilitam uma estratificação de risco. Seguem abaixo os principais subtipos de LLA.

1.1.1.1 Fusão *ETV6-RUNX1*

ETV6-RUNX1, antes conhecido como *TEL-AML1*, é o gene de fusão resultante da translocação entre a banda 13 do braço curto do cromossomo 12 e a banda 22 do braço longo do cromossomo 21. Esta é a mais frequente alteração citogenética encontrada em casos de B-LLA, correspondendo a cerca de 25% deles, e representa um marcador de bom prognóstico (14). Sabe-se também que, por si só,

a t(12;21)(p13;q22) não provoca o aparecimento de leucemia sozinha, de modo que são necessárias mutações adicionais para tanto (13,15).

1.1.1.2 Fusão *BCR-ABL1* (Cromossomo Filadélfia)

A translocação entre o cromossomo 9 e o 22 (q34; q11.2) provoca a fusão dos genes *BCR* e *ABL1*, formando o chamado “cromossomo Filadélfia” e gerando uma proteína quinase quimérica constitutivamente ativada, de modo que sinais de proliferação são liberados constantemente nestas células. Pacientes com esta alteração apresentavam prognóstico ruim, mas com o desenvolvimento dos inibidores de tirosina-quinase esse cenário melhorou bastante (16). Este subtipo ocorre com baixa frequência em crianças, mas contabiliza aproximadamente um quarto dos casos em adultos. Em aproximadamente 80% dos casos de B-ALL *BCR-ABL1+*, o gene *IKZF1* encontra-se deletado (4).

1.1.1.3 Fusão *TCF3-PBX1*

A translocação entre o cromossomo 1 e o 19 (q23;p13) foi inicialmente descrita em 1984 por Carrol e colaboradores como um marcador de prognóstico ruim em pacientes de B-LLA (17). Esta alteração provoca a fusão dos genes *TCF3* (também chamado de *E2A*) e *PBX1*, sendo que o primeiro codifica dois fatores de transcrição, ELSPB1 e TCF3, envolvidos no processo de maturação de células B e o segundo é um gene *homeobox* que aparenta ser expresso em todos os tecidos, menos em células B e T. As consequências desse evento são a expressão não fisiológica de *PBX1*, a trans-ativação de diversos genes e a produção de um fenótipo maligno (18). Apesar da descrição inicial sombria, hoje esta alteração foi quase excluída do rol de estratificação de risco devido a ótima resposta à terapia atual (17).

1.1.1.4 Rearranjo de *MLL*

Os rearranjos de *MLL* (*mixed-lineage leukemia*; também conhecido como KMT2A) são mais incidentes em bebês com menos de 1 ano de idade, conferem um prognóstico sombrio e podem ocorrer em várias formas como t(4;11), t(11;19) e

t(9;11), por exemplo (19). O gene *MLL* situa-se no locus cromossômico 11q23 e codifica uma metiltransferase de lisina, capaz de regular a expressão de outros genes. A concordância de quase 100% em gêmeos monozigóticos indica que essa alteração ocorre ainda no desenvolvimento intrauterino do feto e que, ainda, pode ocorrer metástase transplacentar de um feto para o outro (4). As translocações de *MLL* envolvem vários parceiros diferentes (20) e têm em comum a desregulação de genes pertencentes ao cluster *HoxA*. Na maior parte dos casos, DOT1L é necessária para que essa desregulação ocorra (21,22).

1.1.1.5 Subtipo com Hiperdiploidia e subtipo com Hipodiploidia

Williams e colaboradores identificaram em 1982 o número de cromossomos como um fator importante de predição de prognóstico em LLA (23). Desde então, a hiperdiploidia tem sido reconhecida como um marcador de bom prognóstico e geralmente envolve ganho dos cromossomos 4, 10, 14, 17 e 21 (24,25). Pacientes com hipodiploidia têm prognóstico sombrio e contabilizam aproximadamente 3% dos casos de B-LLA. Essas alterações, no entanto, não provocam sozinhas o aparecimento de leucemia em modelos experimentais, de maneira que hoje está estabelecido que mutações submicroscópicas adicionais são necessárias em caráter cooperativo para o aparecimento da doença e aquisição de resistência a agentes quimioterápicos (5,26).

1.1.1.6 Subtipos de alto risco

1.1.1.6.1 LLA De Precursores Jovens De Células T

As células destas neoplasias exibem um perfil de expressão genética que é semelhante ao de precursores jovens do timo. Coerentemente, essas células cancerígenas acabam por possuir potencial de diferenciar-se em múltiplas linhagens, incluindo a mielóide e a linfóide. No entanto, apresentam um imunofenótipo distinto (CD1a-; CD8-; CD5 fraco; expressão de marcadores de células mielóides e tronco). Genes envolvidos com o desenvolvimento hematopoiético, receptores de citocinas e sinalização RAS, e modificações de

cromatina também se encontram alterados. Cerca de 12% dos pacientes com T-LLA são afetados por este subtipo (3,4).

Os pacientes deste subtipo de T-LLA sofrem de um prognóstico muito pobre, ainda que passando por transplantes, mas alternativas de tratamento como uma quimioterapia mais intensa incluindo asparaginase pegilada e dexametasona, terapias direcionadas a perfis de expressão característicos de células mielóides e terapias epigenéticas podem representar uma melhora neste quadro (3).

1.1.1.6.2 LLA Semelhante a *BCR-ABL1*⁺

Este subtipo é caracterizado por alterações de expressão em *IKZF1* (4), um fenótipo característico do subtipo *BCR-ABL1*⁺. Embora, não expresse a proteína de fusão *BCR-ABL1*, muitos pacientes desta doença exibem vários rearranjos envolvendo tirosina-quinases como *ABL* e *PDGFR* e respondem ao tratamento com Imatinib e Dasatinib. Outros genes que podem apresentar-se afetados por rearranjos são *JAK*, *EPOR*, *CRLF2*, sendo os dois primeiros responsivos à inibição de *JAK* e o último presente em 50% dos casos deste subtipo. O prognóstico para esta doença, infelizmente ainda é pobre (3).

1.1.1.6.3 LLA associada a Síndrome de Down

Pacientes com síndrome de Down apresentam cerca de 20 vezes mais chance de desenvolver LLA. O prognóstico é pobre, uma vez que eles apresentam também maior taxa de recaída e de mortalidade relacionada ao tratamento. Ainda é incerto como o cromossomo 21 extra presente nos indivíduos com síndrome de Down influencia a sua susceptibilidade à LLA, no entanto regiões nos cromossomos sexuais foram identificadas como apresentando deleções submicroscópicas que afetam o gene *CRLF2* (3).

1.2 EPIGENÉTICA E CÂNCER

1.2.1 Epigenética

Aproximadamente todas as células de um organismo possuem o mesmo material genético. Contudo, sabemos que em um mesmo indivíduo coexistem diversas e numerosas linhagens celulares com características bastante distintas e que são especializadas em determinadas funções. Em humanos, podemos citar - a título de exemplo - os hepatócitos e os fibroblastos, que embora tendo a mesma carga genética, são especializados em funções distintas: o primeiro em processamento de nutrientes e substâncias nocivas, e o segundo em formação de tecidos de sustentação. Esses também diferem morfológica e fisiologicamente. Tal diversidade só pode existir devido à possibilidade de expressão diferencial dos genes, isto é, a capacidade das células de ativar a expressão de determinados genes enquanto reprime a de outros (21,27).

A expressão dos genes é dirigida primariamente pela ligação de fatores de transcrição a promotores e *enhancers* relacionados (21). Todavia, a conformação física da cromatina - que pode ser mais condensada (heterocromatina) ou menos condensada (eucromatina) - tem o potencial de tornar tais sequências menos ou mais acessíveis, respectivamente, à ação desses fatores.

Devido ao seu grande tamanho, a molécula de DNA associa-se a proteínas para formar um polímero denominado cromatina, que se dobrando sobre si mesmo, diminui o espaço que ocupa. Este alto grau de compactação no interior de células eucarióticas só pode ser atingido devido à estrutura básica da cromatina ser formada por nucleossomos. Estes são corpos formados pelo enovelamento da dupla-fita ao redor de octâmeros de histonas a uma periodicidade de aproximadamente 147 pares de bases (28).

Histonas são pequenas proteínas altamente conservadas entre eucariotos e que são compostas por um domínio globular e uma cauda N-terminal não estruturada. Existem cinco tipos básicos de histonas: H1, H2A, H2B, H3 e H4.

Na porção interior do nucleossomo, o octâmero de histonas é composto por um tetrâmero de H3 e H4 e dois dímeros H2A-H2B, donde sobressaem as caudas amino-terminais de cada proteína seguindo as fendas menores geradas pela torção

das duplas-hélices em torno de seus próprios eixos. A histona H1 estabiliza o nucleossomo e promove a ligação internucleossomal (28,29).

Devido à sua protrusão para fora do nucleossomo, as caudas N-terminais das histonas estão sujeitas a ação de enzimas capazes de marcá-las com modificações pós-traducionais (PTM's, do inglês *post-translational modification*) (30,31). Sabe-se que estas marcações têm, por diferentes mecanismos, o potencial de facilitar ou diminuir o acesso de fatores de transcrição às regiões promotoras dos genes, de maneira que regulam a expressão destes. Dentre as primeiras alterações epigenéticas descobertas estão alterações moleculares no DNA por fatores específicos, como a inserção de um grupamento metil a um nucleotídeo, que confere diferentes possibilidades de expressão gênica. Tais modificações nos componentes da cromatina, juntas com a compactação dela em eucromatina ou heterocromatina e os níveis disponíveis de fatores de transcrição, estabelecem o cenário para a existência da expressão diferencial dos genes.

Além dos fatores capazes de adicionar modificações, existem aqueles que podem ler essas modificações e outros cuja função é retirá-las. Enzimas como essas são conhecidas como “escritores”, “leitores” e “apagadores”, respectivamente (30). Dado este contexto, sabemos que as alterações epigenéticas são reversíveis e que, portanto, essas enzimas podem ser consideradas como instrumentos para a alteração intencional da expressão gênica (30).

Na literatura, existem numerosos relatos de que os padrões de expressão gerados por essas modificações são herdáveis ao longo das divisões celulares. Assim, foi reconhecida a existência de uma herança informacional, à parte da genética, que não depende de alterações na sequência primária de DNA. A este forma de herança convencionou-se chamar epigenética - etimologicamente “acima da genética” (21,27).

Diversos mecanismos epigenéticos já foram descritos, dentre os quais e além dos já mencionados, podemos citar a expressão de variantes histônicas, a ação de RNA's não codantes e outros.

1.2.1.1 Metilação de DNA

A descoberta de que o DNA pode ser metilado remonta à década de 40 do século passado e a descrição de sua relação com a regulação gênica se deu por meados da década de 80, de maneira que este foi um dos primeiros mecanismos epigenéticos a serem descritos e relacionados com o câncer e permanece até hoje como tendo sido o mais extensivamente estudado (32–35). Em eucariotos, somente resíduos de citosina podem ser metilados, dando origem ao nucleotídeo 5-metilcitosina. Essa reação é catalisada pelas enzimas metiltransferases de DNA (DNMT's) e se dá, geralmente, em resíduos CpG (Citosina-fosfato-Guanina). O genoma humano apresenta sítios com grande concentração destes dinucleotídeos, as chamadas ilhas CpG (CGI's, do inglês *CpG islands*), que podem ser encontradas principalmente em promotores de genes, mas também em corpos gênicos e regiões intergênicas (30,36,37).

A metilação de DNA está envolvida em várias situações naturais de inibição da transcrição como o *imprinting*, a inativação do cromossomo X e o silenciamento de sequências repetitivas (37), de maneira que tem sido vista por muito tempo como uma marcação para a repressão transcricional. Certamente existem muitas evidências de que CGI's metiladas nas regiões promotoras realmente apresentam relação inversa com a transcrição das sequências contidas à jusante, mas outras evidências apontam para um efeito contrário da metilação de CGI's nos corpos de genes, o que tem levado à revisão desta visão em direção a um papel dependente do contexto (30,36). É questionado também se a metilação de DNA não seria um reforço para assegurar o caráter já reprimido, por outros mecanismos, de um segmento gênico (36).

Um modelo aceito para a metilação de DNA consiste em que duas enzimas – DNMT3A e DNMT3B –, em combinação com uma terceira, DNMT3L, sejam responsáveis pela metilação *de novo*, enquanto a enzima DNMT1 é responsável pela marcação de fitas-filhas após a replicação de DNA, com o fim de manter o *status* de metilação. Adicionalmente, sabe-se que a desmetilação, além de poder ocorrer de forma passiva (como já era aceito por muito tempo), também pode ocorrer ativamente pela ação de enzimas desmetilases e glicosilases de DNA (30,36).

Desde a década de 70 do século passado, é conhecido que, além da 5-metilcitosina, existe também a variação 5-hidroximetilcitosina em células de mamíferos. No entanto, até hoje foram descritas ainda mais três variantes: 5-formilcitosina, 5-carboxilcitosina e 3-metilcitosina. As enzimas Ten-Eleven Translocation (TET) foram confirmadas como as responsáveis pela oxidação da 5-metilcitosina e conversão em 5-hidroximetilcitosina e podem estar relacionadas com a produção das outras variantes também. Embora pouco se saiba a respeito do papel de cada uma dessas formas, é provável que sirvam como algum tipo de sinalização para a ligação de determinados fatores, podendo inclusive ter influência na metilação de histonas, outro tipo de mecanismo epigenético (35,38).

A metilação de DNA tem sua importância biológica destacada pelo conhecimento de que ondas de reprogramação geral dos padrões de metilação são realizadas durante a gametogênese e após a fertilização, uma vez que erros nesses processos podem alterar a expressão de genes importantes, como aqueles sob influência de *imprinting*, e trazer consequências sérias para o indivíduo em desenvolvimento (39,40). Adicionalmente, a perda de metilação global (principalmente em sequências repetitivas, conduzindo à instabilidade genômica; e em promotores de proto-oncogenes) e o ganho de metilação em regiões específicas do genoma (principalmente em promotores de genes supressores tumorais) de células é atualmente reconhecida como uma das principais marcas do câncer.

1.2.1.2 Alterações pós-traducionais em histonas

1.2.1.2.1 Acetilação de Histonas

A acetilação de histonas - realizada pelas enzimas acetiltransferases de lisina (KAT's, do inglês *lysine acetyltransferases*) e contrabalanceada pela ação das deacetilases de histona (HDAC's, do inglês *histone deacetylases*) - consiste na adição de uma porção acetil ao grupo ϵ -amino de resíduos de lisina histônicos (41).

As KAT's utilizam como cofator a acetilcoenzima A (Acetil-coA) e são agrupadas em três famílias: GNAT (nominada assim devido aos membros Gcn5 e ScKAT2), MYST (nome devido às iniciais dos membros MOZ, Ybf2/Sas3, Sas2 e Tip60) e p300/CBP. Em contraste, as HDAC's são divididas em cinco classes: I, IIa,

IIb, III (Sirtuínas) e IV, sendo que as classes I, II e IV utilizam cofatores metálicos e produzem acetato e lisina como produtos, enquanto as Sirtuínas fazem uso do cofator dinucleotídeo de nicotinamida e adenina (NAD⁺) e produzem nicotinamida, 2'-O-acetil-ADP-ribose (OAAADPr) e lisina como produtos (31,41).

Já em 1964, Allfrey e colaboradores pensaram na possibilidade deste tipo de alteração influenciar a síntese de RNA (42). Desde então a pesquisa a respeito dessas alterações tem crescido muito devido à constatação de que essas alterações, além de regularem a expressão gênica, são reversíveis e altamente dinâmicas, tendo inclusive relação com o ciclo circadiano e o envelhecimento (43–45). No entanto, este mecanismo não é exclusivo de histonas; a acetilação de proteínas não-histonas tem sido detectada com frequência recentemente. A enzima GCN5, por exemplo, é uma acetiltransferase cujo substrato é a proteína PPARGC1A e o efeito de sua ação é inibitório (32). O envolvimento de acetiltransferases com o câncer também é conhecido. Um exemplo é a observação de que 15% dos casos de ALL com hiperdiploidia têm recaída e a maioria destes casos apresenta mutações no domínio acetiltransferase de CREBBP; e camundongos heterozigotos para uma mutação em CREBBP apresentam alta incidência de tumores hematopoiéticos (47).

Devido ao reconhecimento da relação entre a acetilação de histonas e a carcinogênese, inibidores de acetiltransferases de histonas (HAT's, do inglês *histone acetyltransferases*) e HDAC's têm sido desenvolvidos com o fim de verificar se a inibição destas enzimas poderia ter algum efeito positivo para o tratamento do câncer. Um exemplo é a molécula C646, que possui ação específica sobre p300 e é capaz de suprimir a proliferação de linhagens celulares de melanoma e câncer de pulmão. É interessante observar que existem inibidores naturais de HAT's, entre os quais podemos citar o ácido anacárdico (extraído da castanha do caju), o garcinol (isolado da planta *Garcinia indica*) e a curcumina (obtida a partir do açafrão-da-índia) (46,47).

1.2.1.2.2 Metilação de Histonas

As histonas – proteínas responsáveis pelo empacotamento e enovelamento do DNA – também podem ser metiladas, a partir da adição de um grupamento metil,

oriundo de um doador S-adenosil-L-metionina (SAM), aos seus grupos ϵ -amino presentes em resíduos de lisina e arginina nas caudas N-terminais (41,47–49). Essas alterações estão relacionadas a estados diferenciados de expressão gênica: sabe-se que a metilação de histonas é um mecanismo mais complexo que a acetilação, produzindo diferentes efeitos – repressão ou aumento da expressão – dependendo do resíduo aminoácido e da histona onde ocorre a metilação e do grau de metilação (resíduos de lisina podem ser mono, di ou trimetilados, enquanto resíduos de arginina podem ser mono ou dimetilados). Tal constatação associada a observações referentes a outras modificações pós-traducionais em histonas levaram à formulação da hipótese do código de histonas, que prevê que essas alterações epigenéticas são interpretadas pela célula e definem parâmetros biologicamente importantes em uma maneira análoga ao que ocorre com o código genético (50–52).

Corroborando esta hipótese, observa-se que geralmente a metilação em H3K4, H3K36 e H3K79 estão associadas à ativação transcricional, enquanto a metilação em H3K9, H3K27 e H4K20 associam-se à repressão transcricional (21,30).

As enzimas capazes de adicionar esta marcação são chamadas metiltransferases de histona. No entanto, devido a muitos trabalhos terem demonstrado que essas mesmas enzimas são capazes de metilar outras proteínas não-histonas, existe uma tendência em chamá-las de metiltransferases de proteínas (53).

Quase todas as metiltransferases de lisina possuem um domínio catalítico comum de aproximadamente 130 aminoácidos chamado SET (*S*upressor of *v*ariation, *E*nhancer of *z*este [*E(z)*] e *T*riThorax), que lhes confere a atividade metiltransferase. Excetua-se a família dot/DOT1L, capaz de metilar a lisina 79 da histona 3 e cujos níveis tem sido ligados a manutenção dos telômeros e regiões repetitivas de DNA (54,55).

A família SET de proteínas, encontrada em todos os eucariotos estudados até hoje, pode ser dividida por especificidade de substratos. Assim, temos aquelas proteínas específicas para metilação em H3K4, H3K9, H3K27, H3K36, H3K79 e H4K20, além daquelas proteínas com substratos não-histona, como provavelmente é a SETD7 (48).

1.2.1.3 Outros Mecanismos Epigenéticos

Vários outros tipos de PTM's já foram descritos: a fosforilação, a ubiquitinação, a sumoilação, a biotinilação, a citrulinação, a ADP-ribosilação, a β -N-glicosilação, a isomerização, a crotonilação, a formilação, a propionilação e a butirilação. Além destes, os RNA's não codantes como os miRNA's (que incluem os iRNA's, os shRNA's, os snoRNA's e os piRNA's) e os lncRNA's são discutidos como possíveis fatores epigenéticos. A expressão de variantes histônicas também tem grande importância na fisiologia celular envolvendo processos como a inativação do cromossomo X (30,56,57).

1.2.1.4 Alterações Epigenéticas no Câncer

A visão clássica do câncer é de um conjunto de doenças causadas por aberrações genéticas em genes supressores tumorais ou proto-oncogenes capazes de gerar um fenótipo celular caracterizado por proliferação sem limites, imortalidade celular, evasão da apoptose, a capacidade de extravasar barreiras naturais invadindo tecidos adjacentes e formação de metástases, em um conjunto conhecido como *hallmarks of cancer* (58). Contudo, sabe-se que mecanismos epigenéticos são capazes de regular a expressão gênica e servem como uma camada extra de regulação. Assim, alterações epigenéticas também são capazes de participar no processo de carcinogênese. Já é reconhecido que as células cancerígenas compartilham perda global de metilação do DNA associado ao ganho desta alteração em regiões específicas do genoma (57,59). Ademais, mutações em genes codificadores de proteínas de efeito epigenético também têm sido descritas em associação com o câncer. Relatos envolvendo neoplasias e a família de metiltransferases NSD (*nuclear receptor binding SET domain containing*) têm sido publicados. Localizada na região do genoma hemizigoticamente deletada na Síndrome de Wolf-Hirschhorn, WHSC1 (também conhecido como NSD2 ou MMSET) é uma metiltransferase capaz de metilar a lisina 36 da histona 3 (H3K36) que já foi relacionada à proliferação de mieloma múltiplo e descrita como apresentando expressão diferencial no contexto do câncer, com aumento em tecidos neoplásicos

em relação a suas contrapartes normais. Como mencionado anteriormente, rearranjos envolvendo MLL são responsáveis por um subtipo de LLA de prognóstico muito pobre. No entanto, essa não é a única metiltransferase envolvida com este subtipo. DOT1L é uma proteína que, embora sendo capaz de realizar a adição de um grupamento metil, não possui o domínio SET. Em alguns desses casos de LLA, DOT1L é determinante para a leucemogênese, sendo recrutada pelos parceiros de MLL e necessária para a desregulação da expressão de genes críticos pertencentes ao cluster HoxA (21,60). Efeitos epigenéticos também já foram descritos como metilando proteínas não histonas e regulando a ação destas, a exemplo do que ocorre com o clássico supressor tumoral p53, alvo de metilação por SETD8, SMYD2 e SETD7 (61).

1.2.1.5 Metiltransferase SETD4

O gene *SETD4* encontra-se no locus 21q22.13, transcreve um mRNA de 3001pb e codifica uma proteína metiltransferase de lisina de 440aa. Além do domínio catalítico SET, a enzima contém um domínio Rubisco LSMT *substrate-binding* (Rubs-subst-bind) que provavelmente confere à enzima a capacidade de ligar-se ao substrato (**Figura 1**). Quatro isoformas são preditas pela base de dados *Protein* do Centro Nacional para Informação em Biotecnologia dos EUA (NCBI, do inglês *National Center for Biotechnology Information*) e também quatro são preditas pela base de dados UniProt. No entanto, apenas três parecem ser compartilhadas pelas duas bases.

Faria e colaboradores publicaram em 2013 a primeira caracterização da proteína quanto aos aspectos de estrutura, localização celular, expressão e relação com o câncer de mama (62). Neste trabalho, os autores observaram que em MDA-MB 231 SETD4 é expressa no citosol e no núcleo celular, porém com diferentes padrões de distribuição, de maneira que é possível que ela tenha diferentes alvos em cada compartimento. Além disso, a estrutura de SETD4 é bastante semelhante a de SETD3. No entanto, o principal achado do artigo foi o de que a expressão da metiltransferase estava aumentada em linhagens de câncer de mama triplo-negativas. Ao realizar o *knock-down* da enzima, os autores verificaram uma diminuição na viabilidade de células MDA-MB 231 e na formação de colônias por

elas, que não foi causada por morte celular, mas por meio de uma parada do ciclo celular na fase G1. Quando os níveis de ciclina D1 foram checados, constatou-se que estavam reduzidos nessas células, de maneira que SETD4 pode ter um papel na regulação do ciclo celular, especificamente na transição G1/S; no desenvolvimento do câncer de mama; e talvez de outras neoplasias.

SETD4, assim como SETD3 e SETD6, é classificada como uma metiltransferase de classe VII, ou seja, provavelmente não tem como alvo uma histona, mas outras proteínas (62,63).

Outro dado interessante sobre a proteína vem de um estudo realizado no contexto do carcinoma hepatocelular (HCC, do inglês *hepatocellular carcinoma*), um tipo de câncer conhecido pela sua insensibilidade ao tratamento com o quimioterápico Sorafenib. Ao realizarem uma procura por genes responsáveis pela sensibilidade ao Sorafenib com o uso de iRNA's, Li e colaboradores observaram que o *knock-down* de SETD4 em células da linhagem HepG2 as torna novamente sensíveis ao quimioterápico (64).

Dado o fato de que a literatura a respeito desta proteína é escassa e que os poucos artigos que trazem relatos a envolvendo mostram evidências de que ela possa ser um gene regulador do ciclo celular e da sensibilidade a quimioterápicos, se faz necessária uma caracterização mais profunda de SETD4 com vistas a verificar o seu potencial como alvo terapêutico e questões como a sua relação com a carcinogênese, bem como sua função no tecido normal e seus respectivos substratos demandam uma procura por respostas.

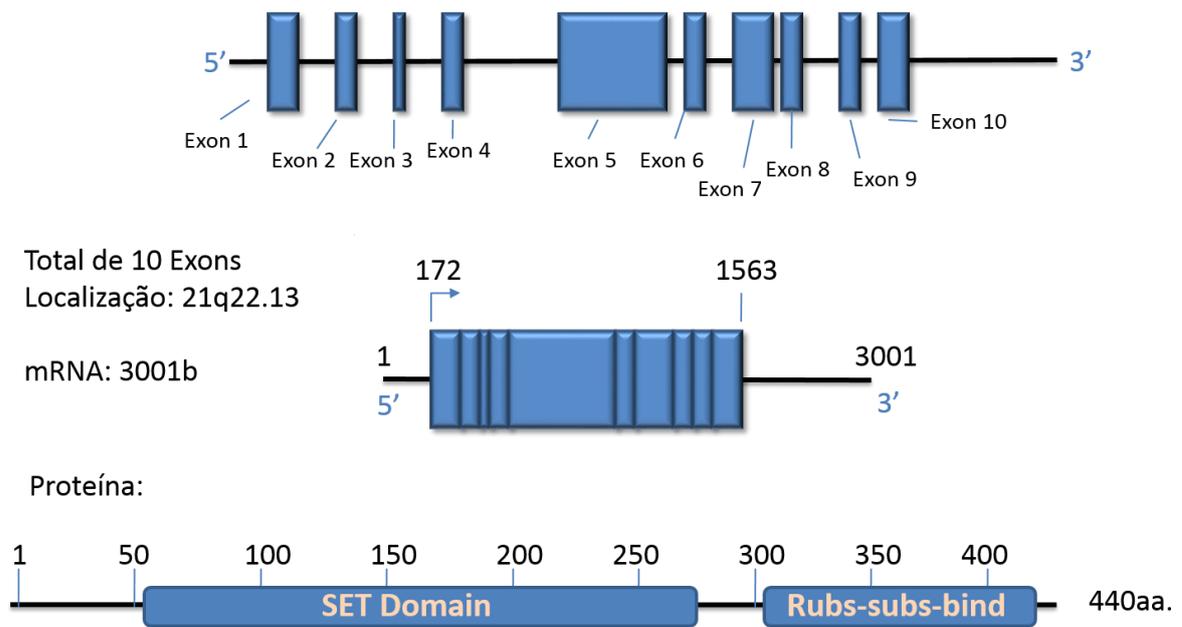


Figura 1. Estrutura do gene, transcrito e proteína SETD4. Na estrutura gênica (superior), os éxons são representados por barras e os íntrons por uma linha preta. Na estrutura do transcrito (meio), os éxons são representados da mesma maneira, porém a linha representa as regiões não traduzidas do primeiro e último éxons. Na estrutura da proteína são representados a cadeia de aminoácidos pela linha preta e os dois domínios proteicos principais: SET e Rubs-sub-bind. Imagem adaptada de Faria e colaboradores (62).

2 OBJETIVOS

2.1 OBJETIVO GERAL

Verificar se os níveis da metiltransferase SETD4 encontram-se diferencialmente expressos em leucemia linfoblástica aguda e, se houver, analisar o impacto dessa alteração na sobrevida de pacientes dessa doença, bem como estudar sua relação com outras enzimas e realizar levantamento dos dados desta proteína em linhagens celulares, tecido normal e tumoral, que estejam disponíveis em bancos de dados online.

2.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos do projeto consistiram em:

1. Avaliar o perfil de transcrição de *SETD4* em amostras clínicas de aspirados de medula óssea de pacientes de LLA por meio de ensaios de expressão diferencial baseados em RT-qPCR do RNA extraído ao diagnóstico e durante o tratamento quimioterápico, e também por meio de análise exploratória de dados de expressão gênica disponíveis em bancos de dados online;
2. Verificar a importância prognóstica de *SETD4* no contexto da LLA através de análise de sobrevida dos pacientes estratificados por categoria de nível da transcrição deste gene;
3. Realizar uma análise exploratória de dados disponíveis referindo-se à metiltransferase SETD4 no contexto de diversas outras neoplasias, disponíveis nos repositórios on-line de bioinformática, que possam complementar os dados do primeiro objetivo, bem como as descrições já realizadas a respeito de SETD4 na literatura.

3 MÉTODOS

3.1 AMOSTRAS CLÍNICAS DE LLA, ENSAIO DE RT-qPCR E ANÁLISE DE SOBREVIVÊNCIA

A obtenção das amostras e dos dados clínicos dos pacientes encontra-se descrita no trabalho de Sakamoto e colaboradores (65,66). Brevemente, aspirados de medula óssea de pacientes com LLA (n = 83) foram obtidos ao diagnóstico, com livre e esclarecido consentimento dos responsáveis (o projeto envolveu apenas crianças), no Hospital da Criança de Brasília José Alencar. O tratamento quimioterápico destes pacientes seguiu o protocolo do Grupo Brasileiro de Tratamento da Leucemias na Infância (GBTLI-93) para aqueles pacientes com B-LLA enquanto o protocolo Berlim-Frankfurt-Munich foi utilizado para os pacientes com T-LLA. Ao 15º e 29º dias de quimioterapia foram obtidos aspirados de medula óssea de 15 desses pacientes. Dados referentes a datas de nascimento, do diagnóstico, do último follow-up e da morte, além de leucometria e percentual de linfoblastos neoplásicos na medula óssea foram registrados (estes dois últimos procedimentos foram realizados com auxílio de uma hematologista consultada). Utilizamos como controle, medula óssea não neoplásica obtida de um grupo de pacientes com púrpura trombocitopênica idiopática (n = 8). O estudo, bem como os formulários de consentimento, foi aprovado pelo Comitê de Ética em Pesquisa da Fundação de Ensino e Pesquisa em Ciências da Saúde (CEP-FEPECS) sob o protocolo 555/11 e número de aprovação no Sistema Nacional de Ética em Pesquisa (SISNEP) 0528.0.013.012-11. O RNA das amostras foi extraído com TRIzol® (Life Technologies) e por meio de RT-qPCR com *primers* específicos para os genes *SETD4* (gene de interesse) e *ACTB* (fator de normalização) foi feita a amplificação do cDNA. A quantificação relativa dos níveis de transcrição de *SETD4* foi feita utilizando-se o método definido pela equação:

$$\text{Eq. 1: } RQ = 2^{-\Delta\Delta Cq}$$

Em que ΔCq corresponde à subtração dos valores de Cq (ciclo de quantificação) das amostras para o gene *ACTB* dos valores de Cq para o gene *SETD4*, enquanto $\Delta\Delta Cq$

corresponde à subtração da média dos ΔCq das amostras não neoplásicas a partir dos valores de ΔCq das amostras de LLA. A análise de sobrevida foi conduzida em software Prism© (GraphPad Software, Inc., La Jolla, CA, EUA) e utilizou o estimador de Kaplan-Meier, o teste de Log-Rank para comparar as curvas e o teste de Mantel-Haenszel para obtenção da taxa de risco (HR, do inglês *hazard ratio*). Com exceção desta última análise, todos os dados referentes a esta coorte foram analisados para significância estatística no software R, considerando-se um $p < 0.05$ como significativo. Os testes utilizados foram o teste não-paramétrico de Mann-Whitney para diferença de médias e o teste de Spearman para testar correlações. Os gráficos também foram gerados no mesmo software.

3.2 ANÁLISE EXPLORATÓRIA DE DADOS DE REPOSITÓRIOS ON-LINE DE BIOINFORMÁTICA

Boxplots resumindo dados de transcrição de *SETD4* em LLA obtidos por microarranjo foram acessados por meio da plataforma OncoMine™ (www.oncomine.org<<http://www.oncomine.org>>, Junho de 2016, Thermo Fisher Scientific, Ann Arbor, MI, EUA) e foram adaptados utilizando-se os pontos máximo e mínimo, e os percentis 10, 25, 50, 75 e 90 disponíveis na versão OncoMine™ Research Edition. Para a busca foram utilizados os seguintes parâmetros de pesquisa: “*SETD4*” (gene), “*cancer vs. normal analysis*” (para tipo de análise) e “*acute lymphoblastic leukemia*” (tipo de câncer). Dados de microarranjo também foram obtidos por meio do banco de dados específico para leucemias Bloodspot (67). Informações sobre a anotação de sondas de microarranjo *Affymetrix* e suas respectivas qualidades foram obtidas no banco GeneAnnot (<http://genecards.weizmann.ac.il/geneannot/>, Instituto de Ciência Weizmann) (68). A plataforma cBioPortal (<http://www.cbioportal.org>) foi acessada (69,70), e as neoplasias que possuíam dados tanto de sobrevida quanto de alterações de número de cópias foram analisadas. Todos os estudos com dados de sobrevida disponíveis no portal em Junho de 2016 foram analisados, exceto o estudo de sarcoma gerado pelo *The Cancer Genome Atlas* (TCGA), pois ainda está em moratória. O teste de *log-rank* foi aplicado para verificar diferenças entre as curvas e o teste de *log-rank* para tendências foi aplicado para verificar a existência de tendência através dos

diferentes escores de alterações de número de cópias (CNA's, do inglês *copy-number alterations*). A plataforma Tumorscape (Broad Institute, Harvard e MIT; <http://www.broadinstitute.org/tumorscape>) foi acessada para obtenção de dados a respeito de picos de amplificação ou deleção envolvendo o gene SETD4. A plataforma Cancer Cell Line Encyclopedia (CCLE; Broad Institute, Harvard e MIT) também foi acessada para obtenção de dados de transcrição gênica em linhagens celulares neoplásicas (71). Todas as análises e gráficos de dados obtidos nestas plataformas foram feitos em software R, exceto os gráficos de sobrevida obtidos através dos dados do cBioPortal, que foram gerados em software PRISM©.

4 RESULTADOS

4.1 EXPERIMENTOS COM AMOSTRAS CLÍNICAS

4.1.1 *SETD4* Apresenta Aumento de Expressão em Amostras de LLA Infantil

Com o fim de obter o perfil de transcrição de *SETD4* em amostras clínicas de aspirados de medula óssea provenientes de pacientes com LLA e de pacientes não neoplásicos, foi realizado um experimento de RT-PCR quantitativa a partir do RNA extraído de tais amostras. Os dados foram comparados entre os grupos após análise por meio do cálculo de quantificações relativas obtidas pela **equação 1** utilizando-se o gene *ACTB* como fator de normalização. A caracterização de nossa coorte pode ser verificada na **Tabela 1**.

O grupo de amostras leucêmicas mostrou maiores níveis de transcrição de *SETD4* em relação ao grupo não neoplásico (teste U de Mann-Whitney, $p < 0.01$) e foi observada uma diferença de 5.48 vezes (**Figura 2**).

Em nossa pesquisa por dados de microarranjo na plataforma OncoMine™ encontramos três estudos envolvendo LLA, sendo que apenas um deles realizava uma comparação entre amostras com este tipo de leucemia e amostras de medula óssea normal ($n = 6$; **Figura 3**). Neste estudo, observamos que independentemente da sonda utilizada, os níveis de *SETD4* mostraram-se elevados em LLA, seja de células B ($n = 85$) ou T ($n = 11$), sendo a diferença observada estatisticamente significativa (todos os valores de P foram menores que 0.001). No segundo estudo foram obtidos valores de intensidade em LLA e em células mononucleares do sangue periférico (PBMC's, do inglês *peripheral blood mononuclear cells*), sendo que duas das três sondas utilizadas apontaram maiores níveis dos transcritos no primeiro grupo. O terceiro estudo avaliou estes mesmos níveis utilizando como controle hematogônias CD10⁺ e CD19⁺, não achando diferença estatisticamente significativa.

Também foi realizada uma pesquisa no banco de dados Bloodspot e foram encontrados dados de transcrição de *SETD4* em amostras de LLA ($n = 750$) e em

amostras normais de medula óssea provenientes de um estudo de microarranjo (72) (n = 73) que, embora presente na plataforma OncoMine™, não tinha dados de medula óssea saudável disponíveis nessa mesma plataforma. Para as três sondas utilizadas, *SETD4* foi detectado com expressão significativamente elevada nas amostras de LLA (P < 0.05; **Figura 4**). O banco *Gene Annot* é mantido pelo Instituto de Ciência Weizmann e, sendo integrado ao banco de informações sobre genes *GeneCards*, provê dados específicos de sondas de microarranjo *Affymetrix* que auxiliam na avaliação da qualidade das anotações das mesmas, de maneira que é possível verificar a possibilidade de erro na associação entre um gene e uma sonda. Os níveis de qualidade variam entre 6 e 1 e seguem os seguintes critérios providos pelo próprio *Gene Annot*:

1. Sondas foram alinhadas a um mRNA, que foi associado a um *GeneCard* via Ensembl ou *Entrez Gene*;
2. Sondas foram alinhadas a um mRNA, que foi associado a um *GeneCard* via *GeneLoc*. Isto é, as coordenadas genômicas do mRNA foram obtidas a partir da Universidade da Califórnia Santa Cruz (UCSC) e o *GeneLoc* foi usado para gerar um link para uma entrada no *GeneCards*, quando pelo menos um exon registrado no *GeneLoc* exibia sobreposição com as coordenadas da UCSC;
3. Sondas foram alinhadas a um mRNA, que atualmente não pode ser associado a um *GeneCard*;
4. Um conjunto de sondas sem um mRNA correspondente. A anotação foi feita baseada no identificador (ID) de *cluster* do *UniGene* (nominalmente aquele cluster que contém a entrada EST – *Expressed Sequence Tag* – de onde o conjunto de sondas foi derivado; somente *clusters* com um título descritivo foram considerados). O ID do *cluster* da *UniGene* foi obtido usando a anotação provida pela *Affymetrix* e não por comparação direta de sequência;
5. Anotação dada pela *Affymetrix* contém uma identificação *Entrez Gene* que foi usada para associar o conjunto de sondas a um *GeneCard*;
6. Anotação original do conjunto de sondas, como provido pela *Affymetrix*.

Interessantemente, embora a sonda 230899_at tenha detectado a menor diferença entre as amostras neoplásicas e aquelas normais, ela é a única das três sondas que exibe nível de qualidade 4 (o terceiro pior) no banco, enquanto as outras duas exibem o mais alto nível de qualidade de anotação.

Alguns fatores clínicos são implicados como determinantes de pior sobrevida entre pacientes de LLA. Para saber se os níveis de SETD4 exibiam alguma influência sobre estes fatores, categorizamos os pacientes em alta expressão ou expressão basal e realizamos o teste de chi-quadrado contra os fatores idade, número de leucócitos, imunologia, e translocações cromossomais recorrentes (**Tabela 2**). Verificamos que para nenhum dos fatores o valor de p foi estatisticamente significativo. No entanto, para o fator número de leucócitos, o valor de p se aproximou bastante de 0.05.

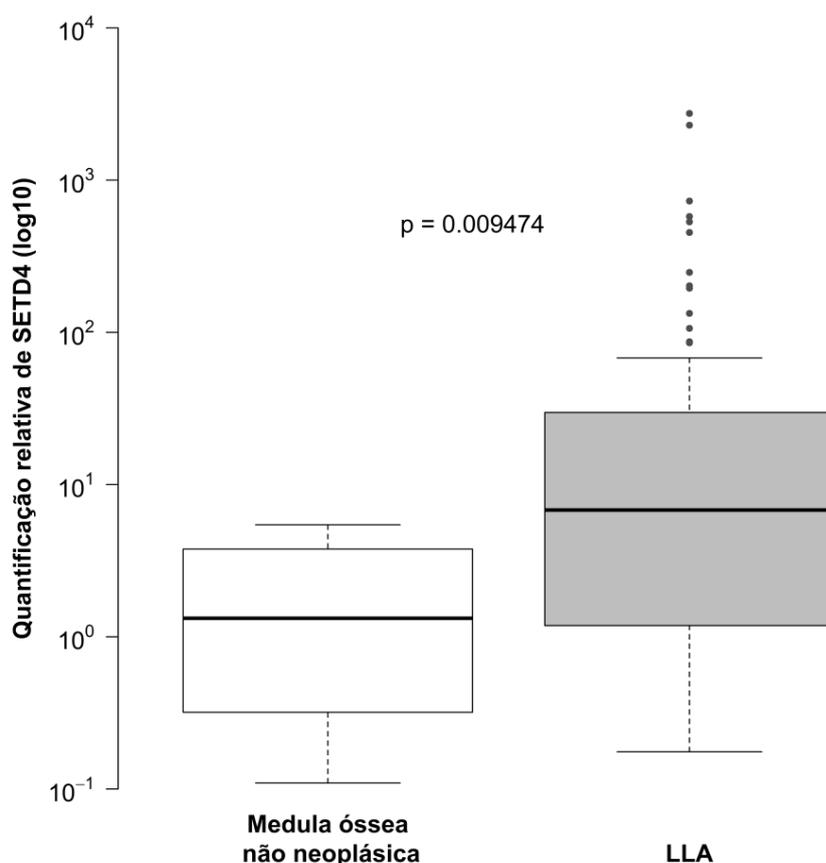


Figura 2. Expressão de mRNA de SETD4 obtida por RT-qPCR em amostras leucêmicas comparadas a amostras de medula óssea não-neoplásicas. Representação em boxplot da quantificação dos níveis de transcrição de SETD4 por RT-PCR quantitativa nas amostras de medula

óssea de pacientes com LLA (n = 83) e de pacientes sem neoplasia (n = 8) que participaram deste estudo. Eixo X, tipo de amostra (LLA ou não neoplásica), Eixo Y, valor da quantificação relativa (RQ) para *SETD4* depois da normalização com *ACTB* (escala logarítmica). Com base na distribuição normal logarítmica dos níveis de expressão de *SETD4*, 13 amostras representadas por círculos foram definidas como *outliers* (círculos preenchidos) de acordo com a regra de rotulamento de *outliers*.

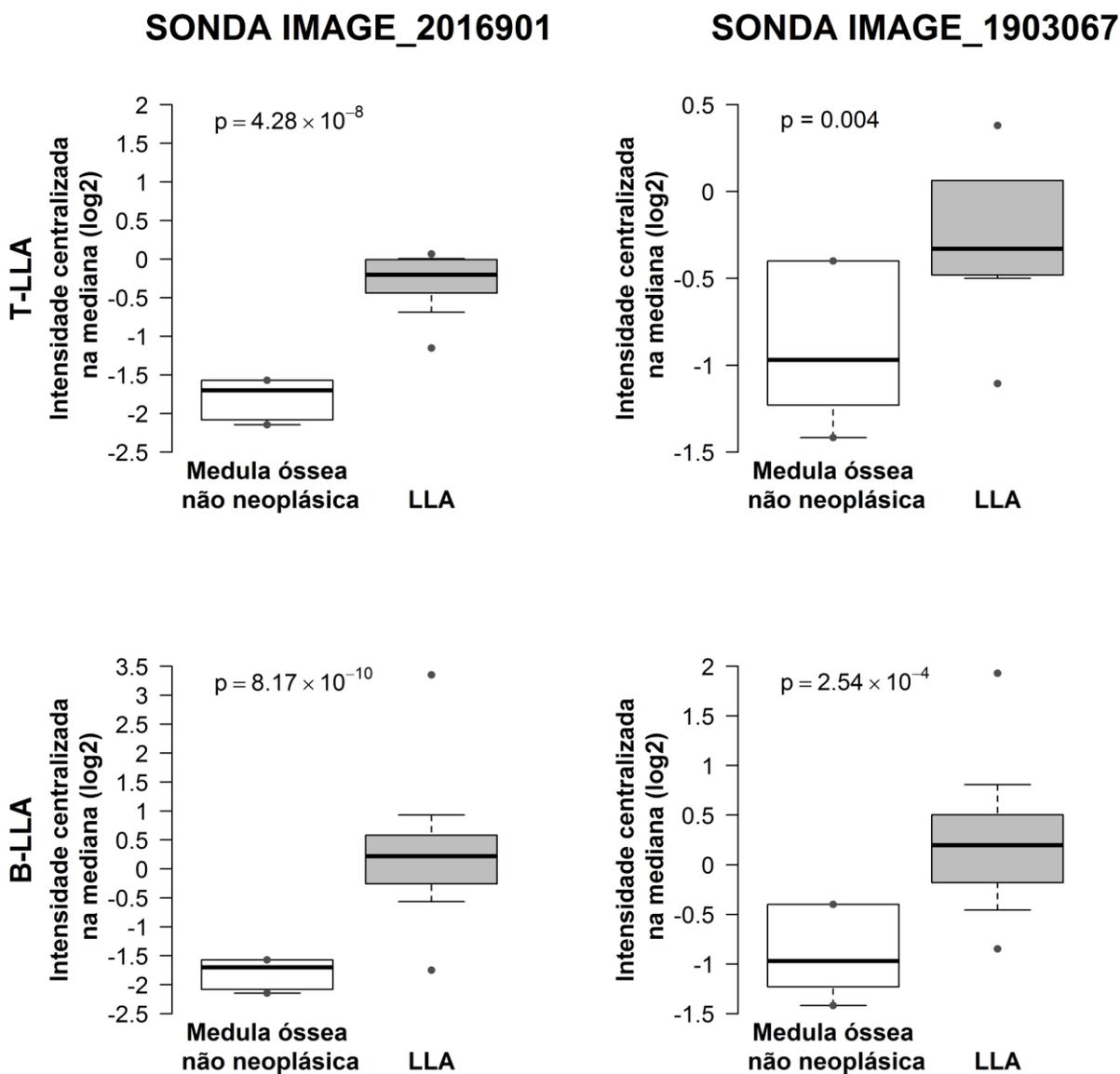


Figura 3. Expressão de mRNA de *SETD4* obtida pela plataforma de dados de microarranjo Oncomine™. Boxplots representando intensidade do sinal obtido por microarranjo com sondas específicas para *SETD4* (IMAGE_2016901 e IMAGE_1903067) a partir da hibridização com mRNA extraído de amostras de medula óssea normal (n = 6) e de LLA. Superior, T-LLA (n = 11), Inferior, B-LLA (n = 85). Dados obtidos pela plataforma OncoMine™.

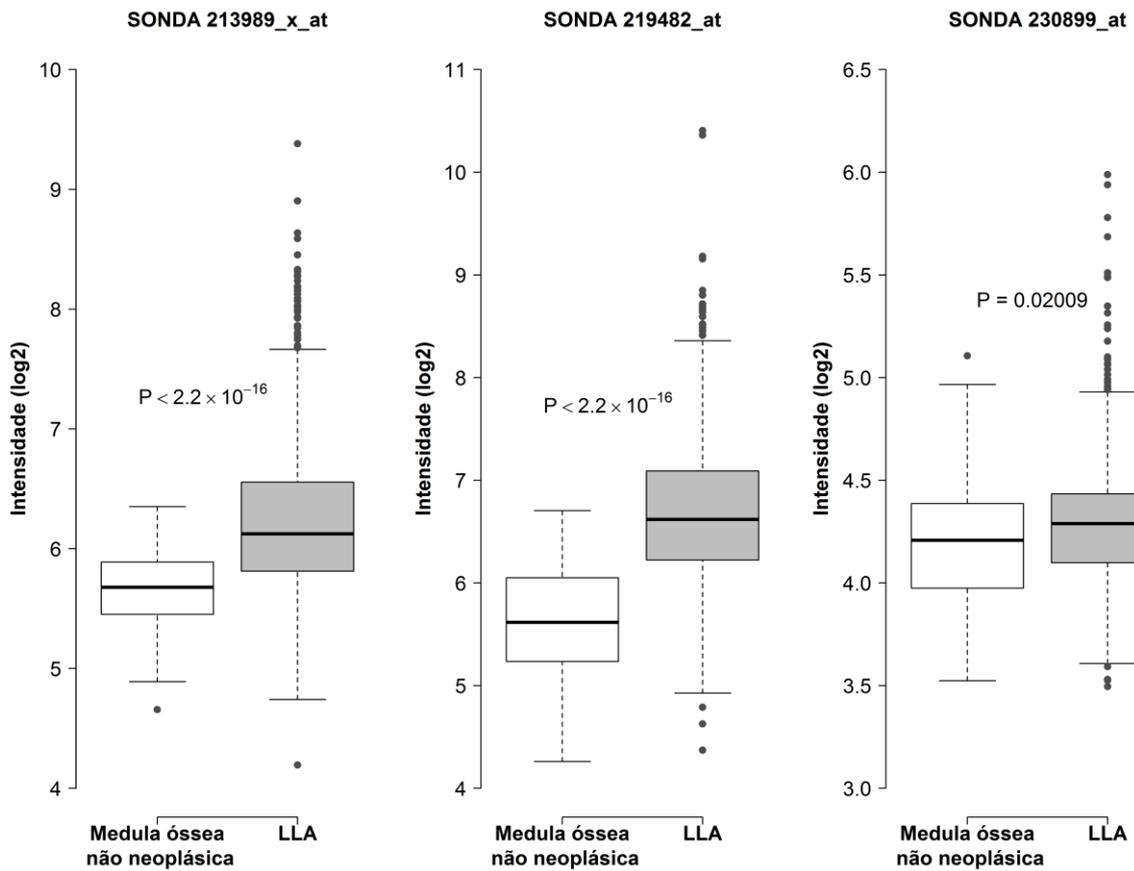


Figura 4. Níveis de transcrição de *SETD4* obtidos por microarranjo em amostras de LLA e de medula óssea normal. Os dados referem-se a níveis de intensidade luminosa detectados por três sondas de microarranjo diferentes e transformados em unidades de Log2. Círculos representam outliers. Cinza, LLA (n = 750); Branco, Medula óssea não maligna (n = 73).

Tabela 1 - Características clínicas da coorte ao diagnóstico e quimioterapia pós-indutória

Características	<i>n</i> (%)
Todos	83
Gênero	
Masculino	43 (51.8%)
Feminino	40 (48.2%)
Idade	
< 1 ano	1 (1.2%)
1-9 ano	53 (63.9%)
> 10 ano	29 (34.9%)
Leucócitos (por μ L)	
< 50000	46 (57.4%)
> 50000	34 (42.5%)
Imunologia	
Pró-B	3 (3.7%)
Pré-B/comum	67 (81.7%)
T-LLA	12 (14.6%)
Grupo de Risco	
Alto	72 (86.7%)
Baixo	11 (13.3%)
Translocações cromossomais recorrentes	
Sem translocações recorrentes	55 (66.3%)
t(12;21)	18 (21.7%)
t(1;19)	8 (9.6%)
t(9;22)	1 (1.2%)
t(4;11)	1 (1.2%)
Status da medula óssea ao D15 da terapia	
< 5% blastos	52 (62.7%)
5-25% blastos	24 (28.9%)
> 25% blastos	7 (8.4%)
Status da medula óssea ao D29 da terapia	
< 5% blastos	77 (93.9%)
5-25% blastos	4 (4.9%)
> 25% blastos	1 (1.2%)

Tabela 2 - Fatores clínicos estratificados por categoria de expressão de SETD4

		Expressão basal	Expressão alta	P-valor (teste χ^2)
Idade	Favorável	38	9	0.1407
	Desfavorável	24	12	
Número de leucócitos	> 50000	17	8	0.0683
	< 50000	38	6	
Imunologia	B	52	19	0.4569
	T	10	2	
Anormalidades cromossômicas recorrentes	Qualquer translocação	22	6	0.5626
	Sem translocação	40	15	

* Análise estatística por teste χ^2

4.1.2 A Expressão de *SETD4* Diminui Durante a Quimioterapia, Bem Como o Número de Células Leucêmicas

Com o fim de verificar se os níveis de transcrição de *SETD4* diminuem com o seguimento da quimioterapia, foram realizados ensaios de RT-PCR quantitativa a partir do RNA extraído de aspirados de medula óssea de 15 pacientes dentre os 83 pacientes de LLA anteriores, retiradas após os dias 15 e 29 do tratamento quimioterápico. Os valores de Cq obtidos foram analisados mediante quantificação relativa utilizando-se a **equação 1** e novamente o gene *ACTB* como fator de normalização.

Dentre estes 15 pacientes, 11 apresentaram redução de expressão de *SETD4* no dia 15 do tratamento quimioterápico em relação aos níveis observados ao diagnóstico. Entre as quatro exceções, encontraram-se os dois únicos pacientes classificados como tendo expressão basal de *SETD4* dentre os 15 analisados. Com exceção do paciente 96 (classificado no grupo de transcrição basal), que apresentou níveis mais altos, todos os outros apresentaram redução da transcrição de *SETD4* no dia 29 em comparação com o diagnóstico (**Figura 5A**).

O efeito esperado de um tratamento quimioterápico é a diminuição do fardo neoplásico, ou seja, da quantidade de células neoplásicas. Sabendo que *SETD4* foi

encontrado mais transcrito em LLA que em células não neoplásicas de medula óssea, a quantidade de células leucêmicas na medula óssea daqueles 15 pacientes foi avaliada também no 15º dia e no 29º dia de quimioterapia com o fim de verificar se poderia ser observada alguma relação entre a transcrição de *SETD4* e o fardo leucêmico. Todos os pacientes analisados mostraram redução no número de células leucêmicas de forma gradual entre o diagnóstico, o dia 15 e o dia 29. No dia 15º dia, 13 dos 15 pacientes (~86.67%) apresentaram número maior que 5% de células leucêmicas, indicando que não houve remissão, enquanto nos pacientes 315 e 322 houve indício de remissão. No entanto, ao 29º dia, o número de pacientes sem indicativo de remissão caiu para apenas 2 (96 e 310; **Figura 5B**). Ao realizarmos o cálculo da correlação de Spearman não foi observada correlação significativa entre os níveis de expressão de *SETD4* e a quantidade de células leucêmicas na medula óssea no 15º dia de indução quimioterápica ($\rho = 0.201$). Entretanto, observamos leve correlação entre a queda na expressão de *SETD4* e a queda no número de células leucêmicas no 29º dia de indução quimioterápica ($\rho = 0.444$; **Figura 5C**). Após esta análise, percebemos que havia um paciente com comportamento discrepante dos demais. Ao trata-lo como um outlier e retirá-lo da análise de correlação, obtivemos uma correlação de 0.427 no dia 15 e de 0.791 no dia 29 (**Figura 5D**).

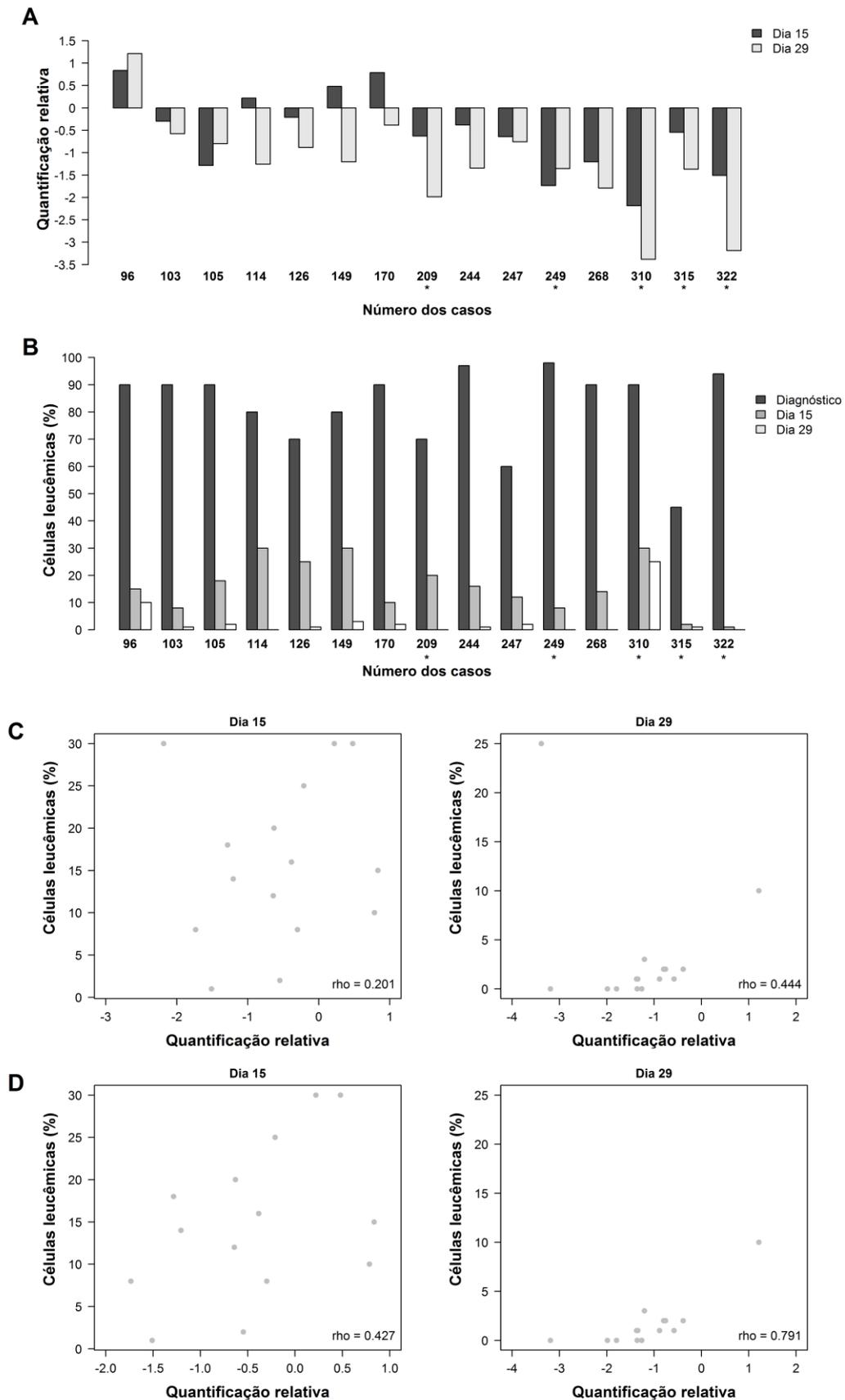


Figura 5. Transcrição de *SETD4* em amostras de pacientes de LLA comparada à diminuição do fardo leucêmico durante a quimioterapia. A, Níveis de transcrição de *SETD4* em 15 pacientes com

LLA nos dias 15 (cinza escuro) e 29 (cinza claro) do tratamento quimioterápico. Os números abaixo das barras referem-se ao registro dos pacientes, asteriscos denotam categorização no grupo de transcrição basal de *SETD4*. B, Percentual de células leucêmicas na medula óssea ao diagnóstico (cinza escuro) e nos dias 15 (cinza claro) e 29 (branco) do tratamento quimioterápico. Os números abaixo das barras referem-se ao registro dos pacientes, asteriscos denotam categorização no grupo de transcrição alta de *SETD4*. C, Gráficos de dispersão mostrando a correlação entre os níveis de transcrição de *SETD4* e a quantidade de células leucêmicas na medula óssea no dia 15 (esquerda) e no dia 29 (direita) do tratamento quimioterápico. D, Gráficos de dispersão anteriores desconsiderando o paciente 310. Coeficientes de correlação de Spearman (ρ) são mostrados no canto inferior direito dos gráficos em C e D.

4.1.3 Os Níveis de Transcrição de *SETD4*, *SETMAR* e *SMYD2* Estão Correlacionados Durante o Tratamento Quimioterápico

Considerando que no estudo anterior desenvolvido por Sakamoto et al. (2014), que utilizou a mesma coorte do presente estudo, foi descrito um perfil de expressão de *SMYD2* em LLA semelhante ao que observamos para *SETD4* e também como um fator de mau prognóstico nessa doença, realizamos a análise da correlação de Spearman para os níveis de transcrição dos dois genes nos 15 pacientes avaliados ao 15^o e 29^o dias de quimioterapia com o fim de verificar se *SETD4* e *SMYD2* poderiam compartilhar alguma relação biológica. Surpreendentemente, uma correlação de 0.925 foi encontrada entre eles nos dois momentos do tratamento quimioterápico (**Figura 7**).

Em estudo preliminar de Sakamoto e colaboradores, foi relatado também aumento substancial na expressão de *SETMAR* em amostras de LLA em relação a controles não leucêmicos (**Figura 6**; dados não publicados). Levando isso em consideração, foi realizada a mesma análise entre *SETD4* e *SETMAR* e foi encontrada também uma alta correlação de 0.87 no dia 15 e de 0.97 no dia 29. Também computamos o nível de correlação entre *SETMAR* e *SMYD2*, que foi de 0.88 no dia 15 e 0.98 no dia 29 (**Figura 7**).

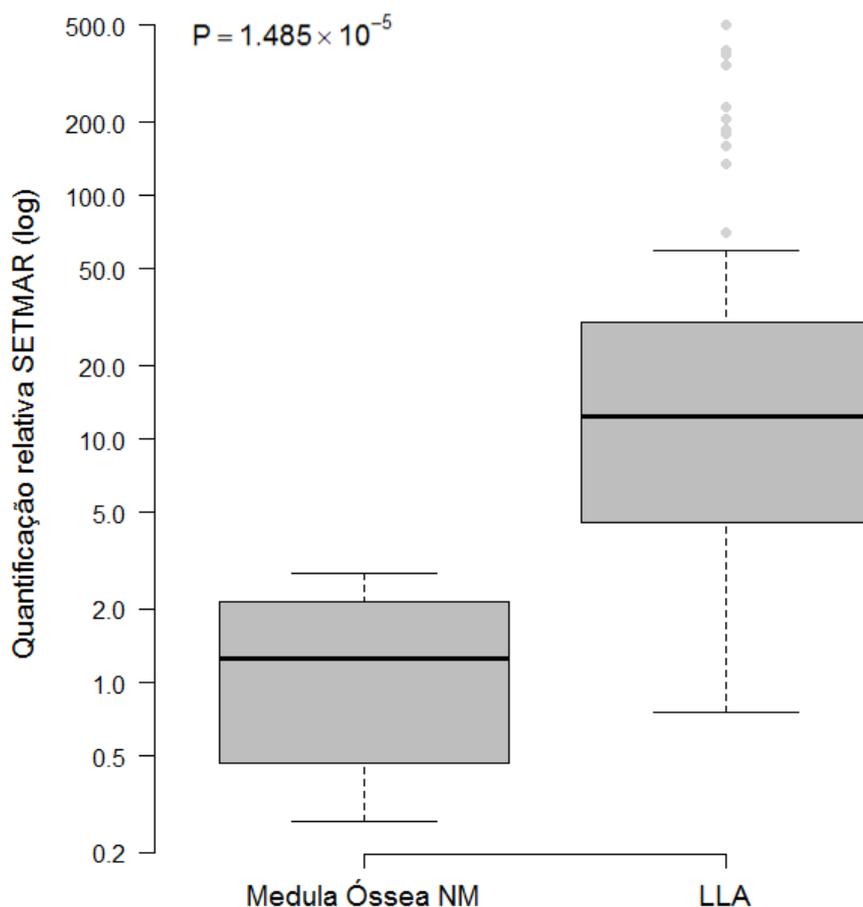


Figura 6. Expressão de mRNA de SETMAR em amostras leucêmicas comparadas a amostras de medula óssea não-neoplásicas. Representação em boxplot da quantificação dos níveis de transcrição de *SETMAR* por RT-PCR quantitativa nas amostras de medula óssea de pacientes com LLA (n = 83) e de pacientes sem neoplasia (n = 8) que participaram deste estudo. Eixo X, tipo de amostra (LLA ou medula óssea não neoplásica), Eixo Y, valor de RQ para *SETMAR* depois da normalização com *ACTB* (escala logarítmica). Com base na distribuição normal logarítmica dos níveis de expressão de *SETMAR*, 12 amostras representadas por círculos foram definidas como *outliers* (círculos preenchidos) de acordo com a regra de rotulamento de *outliers*.

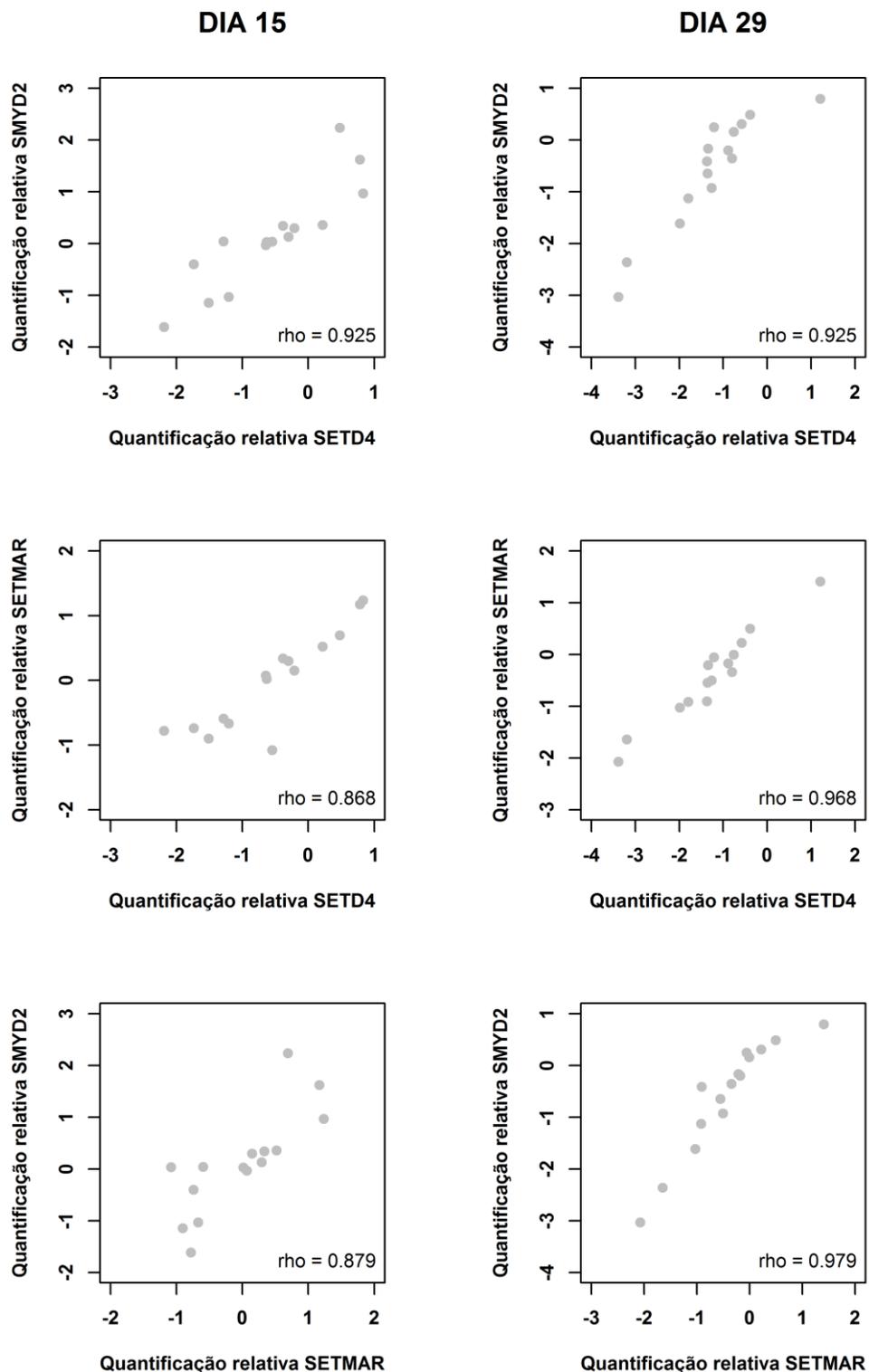


Figura 7. Correlação entre a expressão de *SETD4*, *SMYD2* e *SETMAR* durante o tratamento quimioterápico. Gráficos de dispersão mostrando a correlação existente entre a expressão dos três genes em amostras de medula óssea de 15 pacientes no dia 15 (esquerda) e 29 (direita) da quimioterapia. Todas as correlações foram estatisticamente significativas com $p < 0.01$.

4.1.4 Análise de Sobrevida em Pacientes com LLA

Com o fim de verificar se *SETD4* poderia ser um marcador prognóstico significativo em LLA, foi conduzida uma análise de sobrevida incluindo aqueles mesmos pacientes, levando-se em consideração a categoria de expressão de *SETD4* na qual foram alocados (**Figura 8**). O estimador de Kaplan-Meier foi utilizado para a construção das curvas de sobrevida (uma para o grupo de transcrição basal e outra para o grupo de alta transcrição). O teste log-rank apontou que as curvas são significativamente diferentes ($P = 0.0018$) e o grupo de pacientes com alta transcrição exibiu sobrevida mediana de 610 dias. Não pudemos calcular a sobrevida mediana do outro grupo, uma vez que o número de mortes foi menor que a mediana dos pacientes neste grupo. O tempo de acompanhamento do estudo foi de 1355 dias.

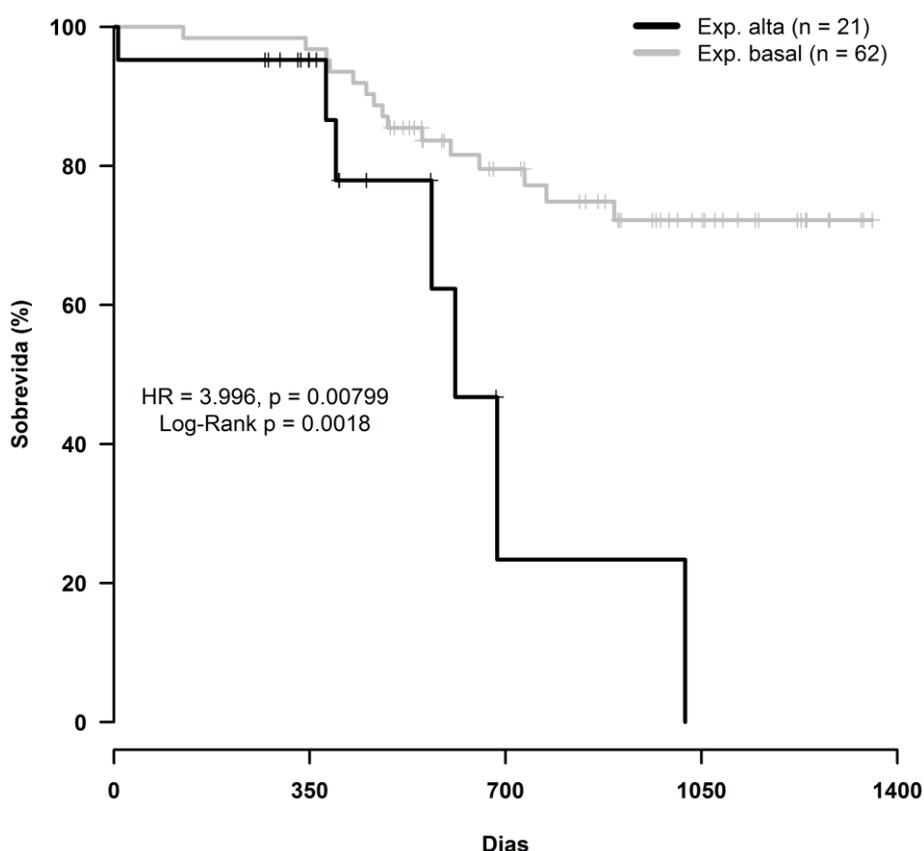


Figura 8. Análise de sobrevida em pacientes de LLA estratificados por categoria de expressão de *SETD4*. Os pacientes foram categorizados em expressão basal (n = 62, linha cinza) ou expressão

alta (n = 21, linha preta) e suas curvas obtidas pelo estimador de Kaplan-Meier são comparadas. HR = riscos proporcionais de Cox.

4.2 ANÁLISES EM BANCOS DE DADOS

4.2.1 Análise de sobrevida com dados do cBioPortal de pacientes estratificados por alterações de número de cópias (CNA's)

Com a finalidade de determinar se a variação na expressão de SETD é um fenômeno comum no agravamento do câncer, foi realizada pesquisa de dados de sobrevida na plataforma cBioPortal para várias neoplasias diferentes. Foram avaliados estudos que incluíam tanto dados de sobrevida como dados de alterações de número de cópias de *SETD4*. Estes compreenderam a dados de pacientes de leucemia mieloide aguda, câncer de bexiga, carcinoma urotelial de bexiga, carcinoma invasivo de mama, carcinoma de esôfago, carcinoma de células escamosas de cabeça e pescoço, adenocarcinoma de pulmão, carcinoma de células escamosas de pulmão, cistadenocarcinoma seroso de ovário e adenocarcinoma de estômago. Foi utilizado o estimador de Kaplan-Meier. Dentre todas as análises realizadas, a que mostrou maior diferença entre os pacientes estratificados por CNA's foi a feita sobre os dados de leucemia mieloide aguda. Neste caso, amplificações ou ganhos de número de cópias de *SETD4* parecem promover o progresso da doença enquanto aqueles pacientes com deleções exibem uma sobrevida um pouco maior (**Figura 9**).

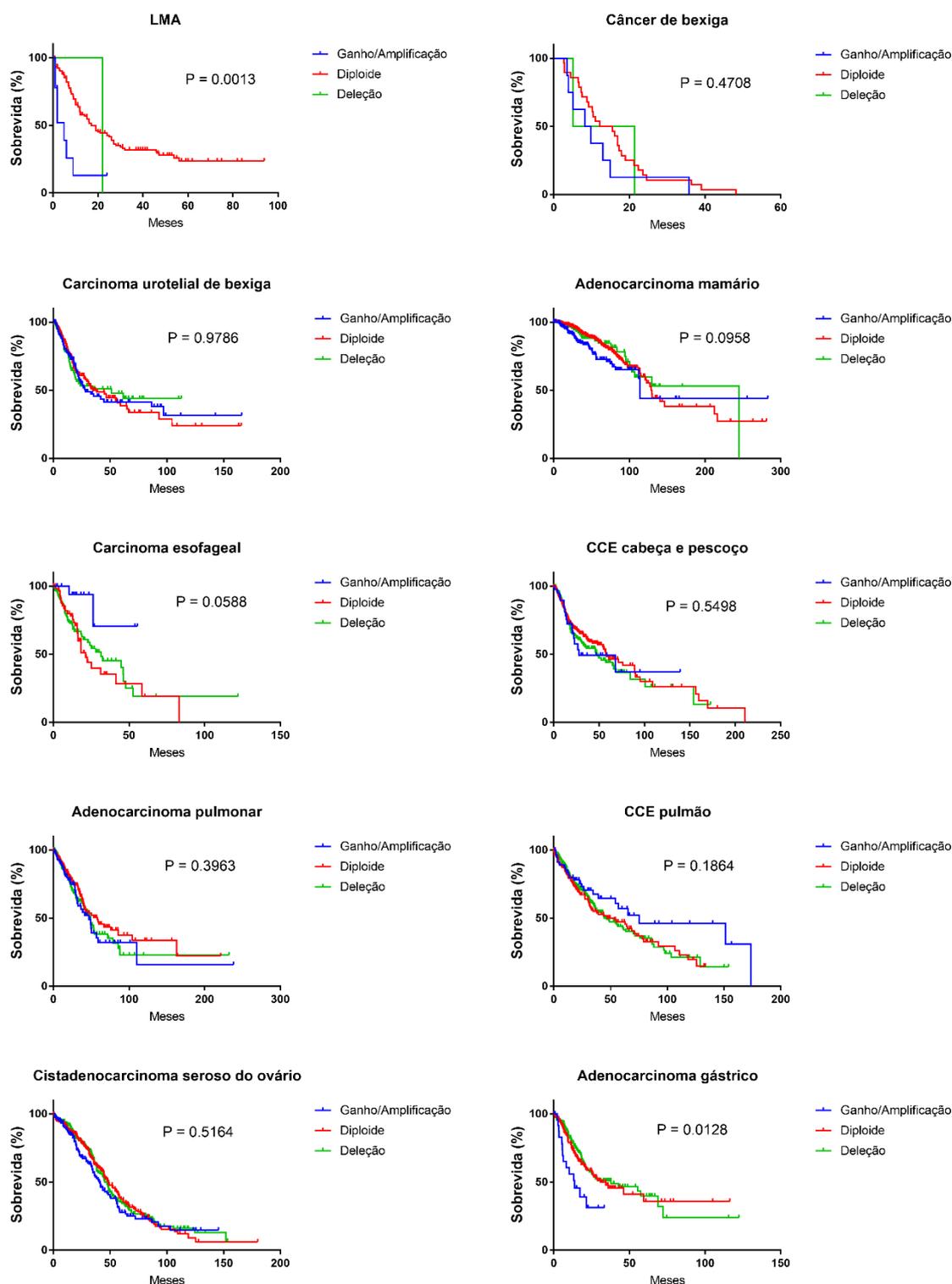


Figura 9. Curvas de sobrevida em pacientes de várias neoplasias estratificados por categoria de CNAs em *SETD4*. A sobrevida foi analisada com o estimador de Kaplan-Meier e os pacientes foram categorizados nos seguintes grupos de CNAs: amplificação/ganho, diplóide (para *SETD4*) e deleção (deleção ou deleção profunda). As curvas foram comparadas com o teste de Log-Rank e os valores de P são mostrados no gráfico. Dados obtidos através do cBioPortal.

4.2.2 Análise de dados de transcrição de *SETD4* no The Human Protein Atlas

A plataforma The Human Protein Atlas contém informações de transcrição (obtidas por RNA-seq) gênica, de expressão proteica (obtidas por imunohistoquímica) e de sublocalização celular (obtidas por microscopia confocal) específicas para cada gene pesquisado e no contexto de diferentes neoplasias e também em tecidos normais (há inclusive uma integração com os dados de RNA-seq de tecidos normais disponíveis na plataforma GTEx Portal, também utilizada neste estudo). Os dados de RNA-seq são normalizados pelo tamanho do gene e pela profundidade da leitura de sequenciamento, obtendo-se valores de FPKM comparáveis entre si (*fragments per kilobase of exon per million fragments mapped*).

Em todos os tecidos analisados, *SETD4* mostrou-se transcrito. No entanto, os maiores níveis de mRNA foram encontrados nos testículos (2.92 vezes maior que a mediana de 6.55 FPKM). Os menores níveis foram detectados no fígado, no músculo esquelético e no pâncreas (23%, 21% e 15% da mediana, respectivamente). Na medula óssea, *SETD4* exibiu transcrição de 53% da mediana enquanto no cólon os níveis foram de 78% da mediana e o reto apresentou uma diferença de 5% a mais que a mediana.

Os níveis de proteína *SETD4* nos tecidos normais foram observados por imunohistoquímica e não mostraram uma relação direta com os níveis de transcrição. A marcação foi mediana nos testículos enquanto houve alta marcação na medula óssea e em células glandulares do cólon. No fígado houve alta marcação de células do duto biliar (nos hepatócitos a marcação foi baixa). Interessantemente, os níveis proteicos de *SETD4* no córtex cerebral também foram altos, embora este seja um tecido marcadamente pouco proliferativo.

Quando os níveis de mRNA de *SETD4* foram comparados aos observados em linhagens celulares de câncer correspondentes, esses se mostraram aumentados em 7 dos 13 casos observados (53%) (**Figura 10**). Interessantemente, quando os transcritos de mRNA dos tecidos linfoides foram estratificados de acordo com o tipo tumoral, *SETD4* evidenciou uma transcrição progressiva, de maneira que apresentou transcrição mediana 1.6 vezes maior em linhagens de linfoma, 2.48 vezes maior em linhagens de mieloma múltiplo e 2.77 vezes maior em linhagens leucemia (**Figura 11**).

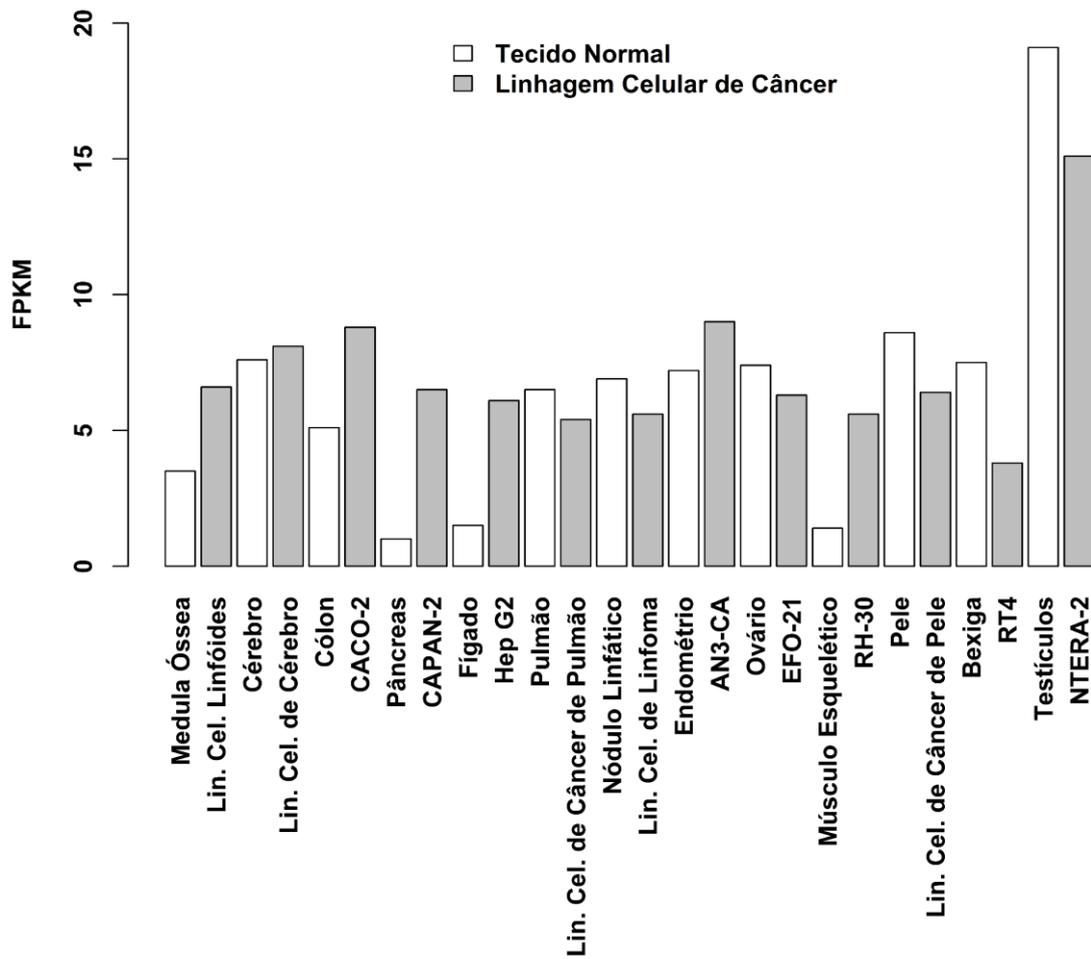


Figura 10. Níveis de mRNA de *SETD4* em tecidos normais e em linhagens celulares de câncer dos respectivos tecidos. Os níveis estão descritos como FPKM (*fragments per kilobase of exon per million fragments mapped*). Tecidos que possuíam mais de uma linhagem correspondente foram comparados à mediana entre as linhagens correspondentes. Branco, tecido normal; Cinza, linhagem celular de câncer.

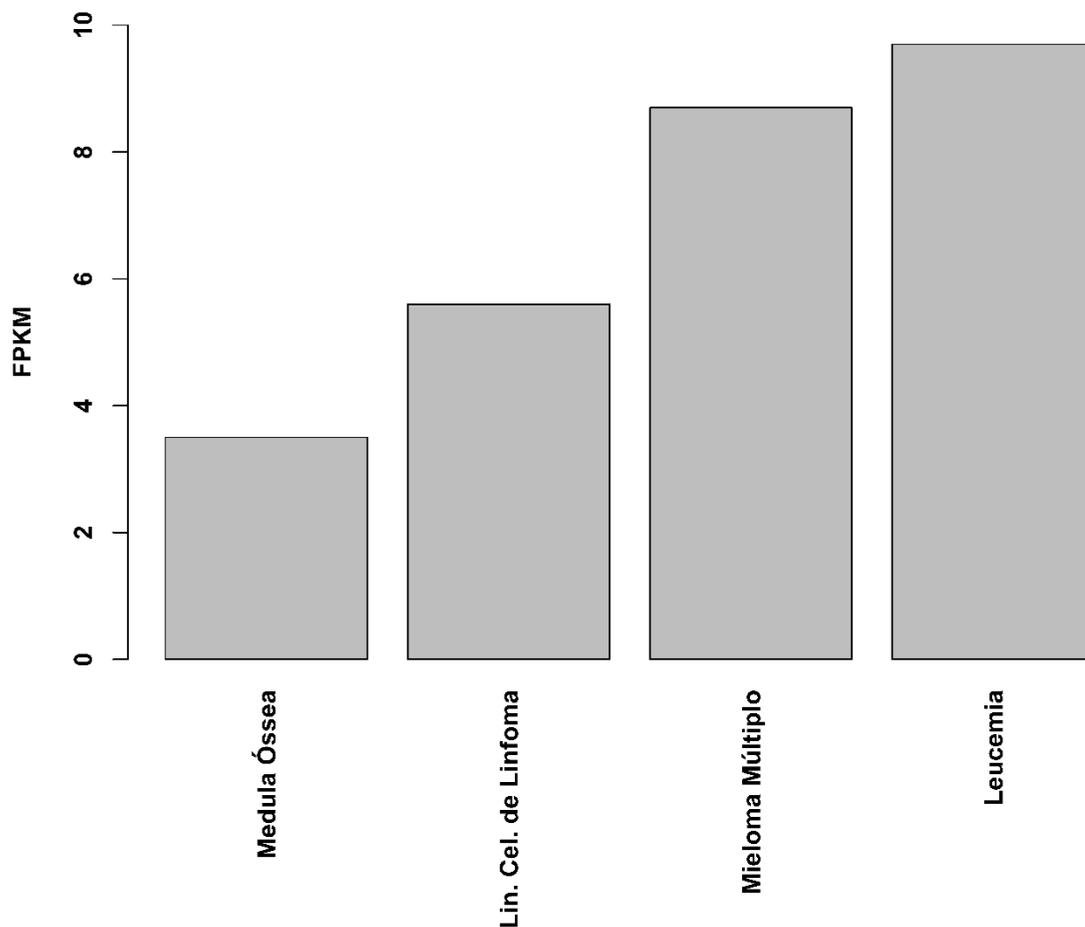


Figura 11. Níveis de transcrição de *SETD4* em medula óssea comparada a linhagens de linfoma, de mieloma múltiplo e de leucemia. *SETD4* encontra-se progressivamente mais transcrito em cada tipo de neoplasia do tecido linfoide. Os níveis estão descritos como FPKM (fragments per kilobase of exon per million fragments mapped).

4.2.3 Análise De Dados No Tumorscape

A plataforma Tumorscape permite a identificação de regiões com incidência de número de cópias maior que aquela esperada como background, os chamados picos de amplificação ou deleção. Também é possível saber a frequência de amostras que possuem alterações focais (abrangendo regiões menores que um braço de cromossomo), globais (abrangem qualquer amplificação ou deleção daquele gene) e de alto nível (abrangem alterações maiores que uma cópia) do gene em questão. Em nenhum dos 14 tipos de câncer disponíveis na base *SETD4*

foi identificado como focalmente amplificado ou deletado. No entanto, quando consideradas todas as neoplasias hematológicas juntas, um pico de amplificação foi encontrado entre as bases 33902218 e 46902240 do cromossomo 21. A frequência global de amplificação para essa região foi de 0.1831, a frequência focal de 0.01 e a frequência de alto nível foi de 0.0558. Apesar disso, um q-valor (taxa de falsas descobertas, a probabilidade do gene apenas sofrer alterações na proporção do background) de 1.0 e a enorme quantidade de genes envolvidos com este mesmo pico (128 genes) diminuem a probabilidade de *SETD4* ser um gene preferencialmente envolvido com estas alterações. Por outro lado, um pico de amplificação em amostras de desordem mieloproliferativa e outro em câncer de mama foram identificados. O primeiro corresponde à região entre as bases 31922227 e 37522096 e o segundo entre as bases 14032696 e 46902240, ambas pertencentes ao cromossomo 21. No entanto, pelas mesmas razões anteriormente citadas, é improvável que *SETD4* seja um gene preferencialmente amplificado nessas regiões (**Tabela 3**).

Tabela 3 - Informações sobre picos de amplificação/deleção abrangendo *SETD4*

Tipo de câncer	Pico mais próximo	# Genes no pico	Q-valor	FGA	FFA	AAN	FGD	FFD	DAN
Todos os hematológicos	Chr21:	128	1.0	0.18	0.01	0.06	-	-	-
Desordem mieloproliferativa	Chr21:	46	0.352	-	-	-	0.00	0.00	0.0
Câncer de Mama	Chr21:	209	1.0	-	-	-	0.16	0.02	0.0

* FGA = frequência global de amplificações; FFA = frequência focal de amplificações; AAN = frequência de amplificações de alto nível; FGD = frequência global de deleções; FFD = frequência focal de deleções; DAN = frequência de deleções de alto nível

4.2.4 Análise de Dados de Expressão na Cancer Cell Line Encyclopedia (CCLE)

O Broad Institute (Harvard e MIT) mantém o portal CCLE como uma base de dados de expressão gênica gerados por microarranjo de amostras de linhagens celulares. Sendo assim, informações a respeito destes modelos para várias neoplasias diferentes encontram-se disponíveis. Nesta pesquisa foi verificado que dentre todas as linhagens neoplásicas, aquelas que pertencem ao grupo de doenças

hematológicas parecem exibir altos níveis de expressão de *SETD4* quando comparadas às outras (**Figura 12**). A maior transcrição foi detectada em trombocitemia essencial e a menor em oligodendrogloma.

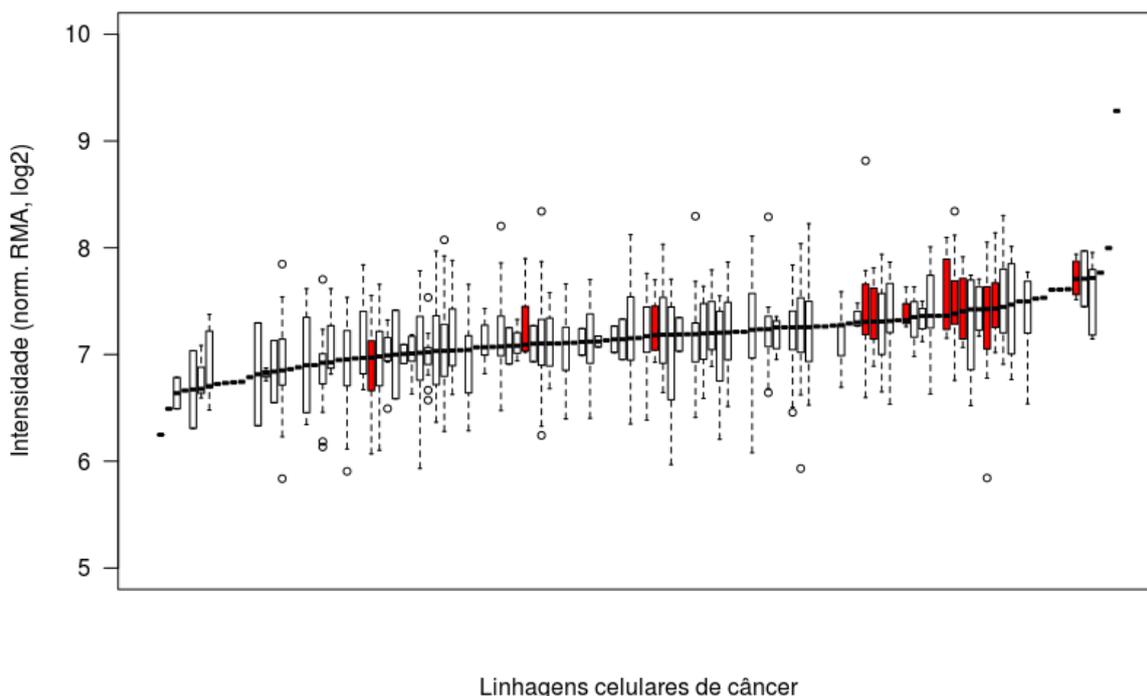


Figura 12. Expressão de *SETD4* em amostras de várias neoplasias obtidos na *Cancer Cell Line Encyclopedia*. Os dados foram obtidos por microarranjo e mostram as intensidades transformadas em escala log₂ e normalizadas pelo método robust multi-array average (RMA). A ordenação foi feita pela mediana. Linhagens de leucemia e de linfoma estão destacadas em vermelho.

4.2.5 Análise de Dados de Expressão no GTEx Portal

A plataforma *Genotype-Tissue Expression Portal* (GTEx Portal) dispõe de dados de transcrição gênica obtidos por RNA-seq *post mortem* em diversos tecidos de corpos humanos e normalizados pelo tamanho do gene e depois pela profundidade da leitura de sequenciamento para obtenção de valores de RPKM (*reads per kilobase of transcript per million mapped reads*). A análise de expressão diferencial obtida para *SETD4* mostra que o tecido que exibe menor quantidade de

transcritos é o sanguíneo enquanto o tecido que exibiu maior nível de transcritos foram os testículos (**Figura 13**). A diferença mediana entre os dois foi de 11.121 RPKM. A tabela contendo o número amostral para cada tecido encontra-se no **anexo I**.

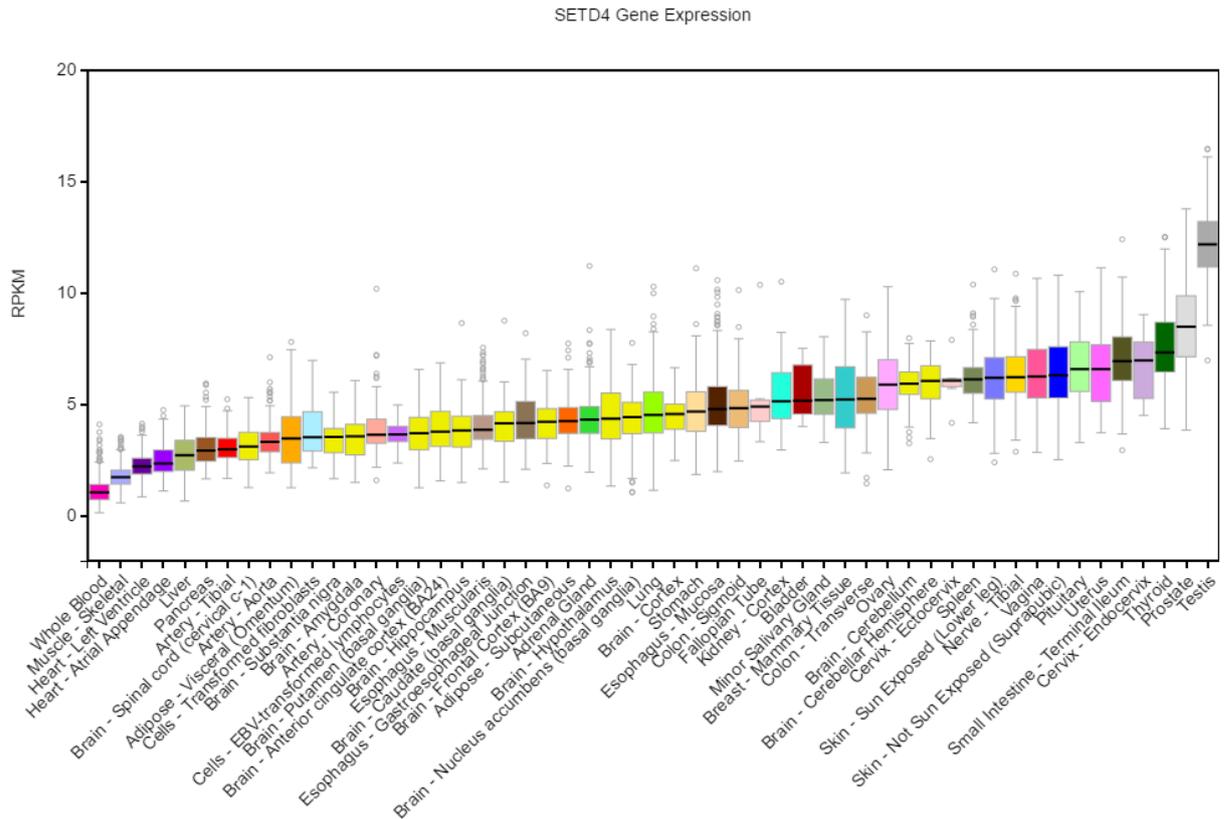


Figura 13. Expressão de *SETD4* por RNA-seq em diversos tecidos não neoplásicos humanos. Os valores correspondem a medidas de RPKM e os tecidos estão ordenados por valor da mediana, sendo cada um codificado por cores individuais.

4.2.6 Análise de mutações na plataforma COSMIC e cBioPortal

A plataforma *Catalogue Of Somatic Mutations in Cancer* (COSMIC) oferece dados processados de mutações abrangendo vários tipos de tumores. Em relação a todas as amostras de tumores disponíveis, a maior proporção de mutações que ocorrem em *SETD4*, representando mais da metade dessa população de dados, é de substituição com troca de sentido e, junto com as substituições sinônimas e substituições sem sentido, contabiliza mais de 90% de todas as mutações

registradas neste gene (**Figura 14 e Tabela 4**). Seis grupos de dados específicos de mutações no contexto da LLA foram encontrados nessa plataforma. No entanto, apenas um deles, específico de B-LLA, possuía amostras com mutações em *SETD4*, sendo a proporção destas de 0.9% em relação ao total de amostras (1 amostra entre o total de 110). A mutação identificada, neste caso, foi p.L425L (c.1275C>T). Na plataforma cBioPortal constavam registradas 88 mutações de substituição e 9 mutações truncantes (stop códons prematuros) entre as diversas neoplasias disponíveis. A mutação mais frequente foi a R14Q, que ocorre fora dos principais domínios da proteína, seguida da R90Q, que ocorre dentro do domínio SET (**Figura 15**). Também foram analisados os dados de previsão de impacto das mutações na função da proteína, obtidos via *Mutation Assessor*, também no cBioPortal. (**Tabela 5**). A maior parte das mutações foi descrita no domínio SET e dentre estas a maioria foi avaliada como de médio impacto.

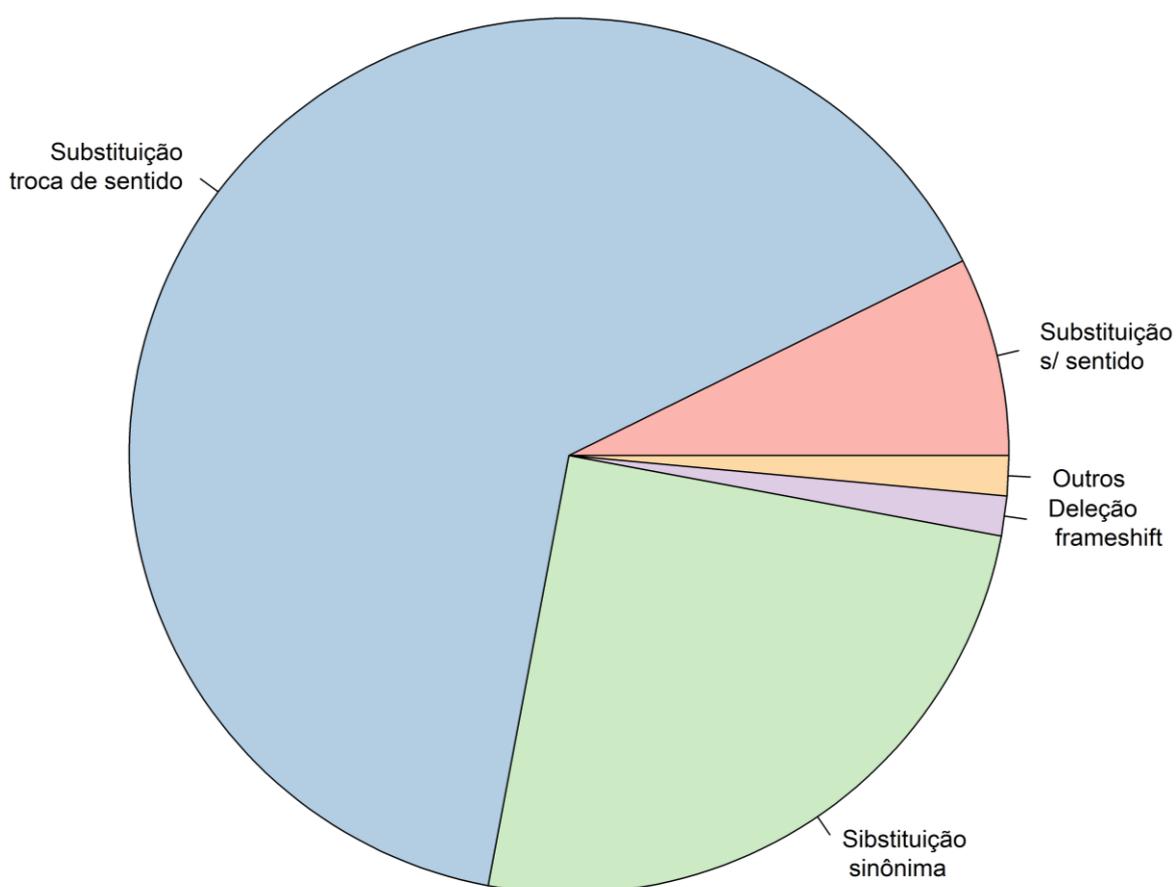


Figura 14. Gráfico de pizza ilustrando a proporção das frequências de cada mutação de *SETD4*. Tipos de mutações estão descritos na legenda juntamente com as cores que os representam.

Tabela 4 - Mutações de *SETD4* registradas na plataforma COSMIC

Tipo de mutação	Número de amostras	Proporção (%)
Substituição sem sentido	5	7.14
Substituição troca sentido	44	62.86
Substituição sinônima	17	24.29
Inserção inframe	0	0.00
Inserção frameshift	0	0.00
Deleção inframe	0	0.00
Deleção frameshift	1	1.43
Complexa	0	0.00
Outro	1	1.43
Total	70	100

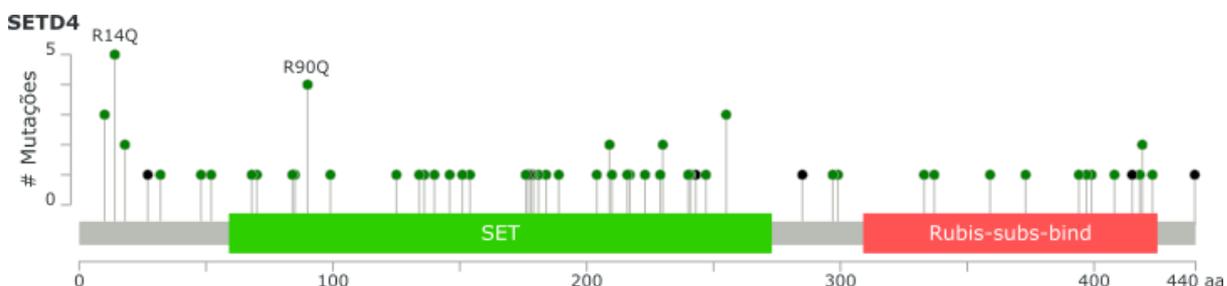


Figura 15. Diagrama de mutações em SETD4 em amostras de vários tipos de câncer obtido pela plataforma cBioPortal. A cadeia proteica de SETD4 é representada na porção inferior com os domínios SET e Rubs-subst-bind destacados em verde e vermelho, respectivamente. Círculos indicam mutações sobre o local onde ocorrem, eixos indicam a o número registrado das mesmas. Em destaque estão as mutações com primeira e a segunda maior frequência, causando a troca de aminoácidos R14Q e R90Q respectivamente. Verde escuro, mutação com troca de sentido; preto, mutação truncante.

Tabela 5 – Predição de impacto de mutações *missense* na função da proteína SETD4 com informações geradas pelo *Mutation Assessor* (cBioPortal)

Impacto	Domínio SET	Domínio Rubs-subst-bind	Outras regiões	Total
Baixo	17	9	11	37
Médio	27	5	4	36
Neutro	8	2	4	14
Total	52	16	19	87

5 DISCUSSÃO

Dado o crescente número de publicações relacionando a maquinaria epigenética com o desenvolvimento do câncer e, em especial, das metiltransferases de lisina, neste trabalho visamos caracterizar os perfis de transcrição de *SETD4* e o seu impacto na carcinogênese, assim como o panorama de mutações, em vários tumores, dando um enfoque especial em leucemia linfoblástica aguda.

O primeiro trabalho a realizar uma caracterização do comportamento de *SETD4* em um tipo de câncer foi o de Faria e colaboradores em 2013. Os autores verificaram que esta proteína se encontra com expressão aumentada em células de câncer de mama triplo-negativo e que a diminuição da expressão dessa proteína provoca redução da proliferação da linhagem celular MDA-MB 231 (62). Tais resultados apontam para um possível papel de *SETD4* como regulador da proliferação de câncer de mama triplo-negativo. No presente trabalho apresentamos uma caracterização do perfil de transcrição de *SETD4* em outro tipo de neoplasia, a leucemia linfoblástica aguda, além da avaliação da sua importância na leucemogênese por meio do acompanhamento dos níveis de transcrição desta enzima durante a quimioterapia de pacientes com esta enfermidade. Foi verificado que, ao diagnóstico, os níveis de transcrição de *SETD4* apresentaram aumento mediano de 5.4 vezes em pacientes de LLA quando comparados aos de pacientes controle sem leucemia. Este dado foi corroborado pela pesquisa no banco de dados OncoMine™, que mostrou um aumento de *SETD4* em amostras de LLA quando comparadas a medula óssea normal e, também a células mononucleares do sangue periférico e hematogônias. Finalmente, os dados obtidos na plataforma Bloodspot confirmam mais ainda essa observação. É interessante o fato de que mesmo com a divisão das amostras pelo subtipo T ou B da doença, a expressão da proteína permanece aumentada em relação ao grupo não neoplásico, o que foi observado tanto nos dados produzidos por nosso laboratório (dado não mostrado) quanto no OncoMine™. Sakamoto e colaboradores, utilizando a mesma coorte do presente trabalho, relataram significativo aumento de *SMYD2* nos pacientes com LLA em relação aos controles não leucêmicos. Diferença similar foi observada para *SETD4*,

mas com uma correlação de expressão mais significativa ligada à resposta ao tratamento, evidenciando que *SETD4* representa um novo marcador molecular significativo para o acompanhamento da doença. O fato deste aumento de expressão ter sido observado em câncer de mama e, agora, em LLA, levantou a hipótese de *SETD4* ser diferencialmente expresso em outros tipos tumorais e, conseqüentemente, um gene com um possível papel comum no processo de carcinogênese em várias neoplasias. Os dados obtidos na plataforma *The Human Protein Atlas* mostram a expressão de *SETD4* em linhagens celulares comparada com a observada em tecidos normais. Na maioria das linhagens, a enzima exibiu níveis de transcrição maiores que as contrapartes teciduais, inclusive no caso da comparação entre o tecido de medula óssea com linhagens de células linfoides. É possível que *SETD4* seja uma proteína com importância na carcinogênese de várias neoplasias, e parece que este é o caso especialmente em neoplasias hematológicas, como indicam os resultados obtidos na *Cancer Cell Line Encyclopedia* em conjunto com os dados de sobrevivência para leucemia mieloide aguda obtidos na plataforma cBioPortal. Interessantemente, este não parece ser o caso para o linfoma, uma vez que a comparação de dados do *The Human Protein Atlas* entre nódulos linfáticos normais contra linhagens de linfoma, mostrou haver menor transcrição de *SETD4* nos primeiros. É interessante também a observação do caráter progressivo entre as transcrições dessa metiltransferase em medula óssea normal e linhagens de linfoma, mieloma múltiplo e leucemia. No entanto, tais observações carecem de confirmação por técnicas mais específicas e robustas como a RT-qPCR, uma vez que foram obtidas por meio de RNA-seq e microarranjo, que são técnicas de alto rendimento, mas que não são tão precisas quanto essa.

Com a finalidade de se verificar o impacto da expressão diferencial de *SETD4* na leucemia, a transcrição dessa enzima, bem como o percentual de células malignas na medula óssea, foram avaliados ao diagnóstico e em dois momentos durante o tratamento quimioterápico dos pacientes de LLA. Embora uma correlação significativa entre a queda do fardo leucêmico e a transcrição de *SETD4* não tenha ficado clara a princípio, foi perceptível que durante o tratamento, os níveis de transcritos da proteína sofreram queda, o que parece refletir a maior transcrição observada nas células leucêmicas. Além disso, o valor da correlação aumentou do dia 15 para o dia 29 e percebe-se a diminuição da dispersão dos dados.

Percebemos que o paciente 310 representava um possível outlier e refizemos a análise desconsiderando-o, de maneira que obtivemos um índice de correlação muito maior. Devido ao pequeno número amostral utilizado nesta pesquisa, é necessário que um estudo de maior porte confirme esta tendência. Resultado semelhante foi observado por Sakamoto e colaboradores para o gene *SMYD2*, que também teve a transcrição reduzida durante quimioterapia, porém com um coeficiente de correlação ligeiramente maior em relação ao fardo leucêmico ($R^2 = 0.5$) (66). Por outro lado, durante este processo, *SETD4* exibiu uma correlação alta com a transcrição do *SMYD2*, sugerindo um possível regulador transcricional comum. Muitos cânceres apresentam transcrição elevada de *SMYD2* e esta proteína já foi relacionada inclusive com a agressividade de câncer de estômago e taxas de recorrência nesta doença (66,73–75). Já foi descrito também que *SMYD2* pode metilar p53 e RB1 (76–78), sendo que quando ele realiza a metilação desta última, E2F (um alvo de RB1) fica livre para promover a passagem da fase G1 para a fase S do ciclo celular (79). Interessantemente, esta é a mesma transição sobre a qual *SETD4* exerceria influência no câncer de mama segundo Faria e colaboradores (62). Assim, é possível que *SETD4* e *SMYD2* sejam regulados pelo mesmo fator transcricional, mas também é possível que venha a interferir juntamente sobre o mesmo fenômeno celular. *SETMAR*, por sua vez, é uma enzima que possui um domínio SET, mas também um domínio Mariner transposase. Ela tem sido relacionada ao mecanismo de reparo de junções não homólogas (*non-homologous end junction*, NHEJ) e à prevenção de translocações cromossômicas (81). No entanto, pouco se sabe ainda a respeito de sua importância no câncer e na leucemogênese, em especial. Esta observação da desregulação de *SETMAR* em LLA é de bastante importância e indica a necessidade de realização de estudos futuros mais aprofundados sobre a relação desta enzima com a leucemogênese de LLA. É interessante que *SETMAR* também tenha exibido correlação alta com *SMYD2*.

Neste trabalho também foi avaliado o impacto de alterações de *SETD4* na sobrevida de pacientes de câncer. Foi verificado que, em LLA, a transcrição alta dessa enzima coincide com um prognóstico pior que o dos demais pacientes. A pesquisa na plataforma cBioPortal também revelou que, em outros tipos de câncer, ampliações e ganhos de número de cópias de *SETD4* parecem influenciar a

sobrevida de pacientes com neoplasias. Em leucemia mieloide aguda, em adenocarcinoma de estômago e carcinoma de pulmão, ampliações e ganhos de SETD4 parecem ter impacto profundo sobre o prognóstico dos pacientes. Embora SETD4 não esteja focalmente amplificado ou deletado em nenhuma neoplasia com dados disponíveis na plataforma Tumorscape e as CNA's afetando este gene não sejam frequentes, nossos dados mostram que essa enzima representa um marcador de pior prognóstico, a exemplo do que ocorre com as translocações de MLL, casos pouco frequentes, mas que infelizmente tornam o prognóstico dos pacientes muito mais sombrio.

O espectro mutacional de SETD4 abrangendo diversas neoplasias, incluindo leucemias e tumores sólidos, foi avaliado em duas plataformas, a saber COSMIC e cBioPortal. Mutações de substituição com mudança de sentido responderam pela maior proporção entre todas as identificadas. Muitas dessas ocorreram no domínio SET, responsável por conferir a atividade metiltransferase para a proteína, e inclusive têm um impacto médio predito pelo *Mutation Assistant*. Pode-se especular que tais mutações venham a prejudicar a função enzimática normal de SETD4 no contexto do câncer, de maneira que seria possível uma desregulação na expressão daqueles genes alvo desta proteína. O domínio Rubs-sub-bind, que é responsável pela capacidade de muitas proteínas se ligarem ao seu substrato (62,80), sofre com algumas mutações de substituição com troca de sentido também. Assim, é possível que nesses casos, a enzima não consiga aderir ao seu alvo, prejudicando sua capacidade enzimática. Algumas mutações truncantes, ou seja, códons de parada prematuros, foram identificadas fora dos dois domínios principais, mas também há um registro dentro do domínio Rubs-sub-bind. Em qualquer caso, uma proteína incompleta é gerada com alta probabilidade de perda de função. Devido a estes possíveis resultados das mutações e seu potencial de influenciar a progressão da doença, será muito interessante que novos estudos verifiquem qual seja a função de SETD4 em tecidos normais e avaliem quais sejam os impactos de cada uma dessas mutações, principalmente as mais frequentes (R14Q e R90Q), na função enzimática da proteína e como elas poderiam contribuir para a carcinogênese.

6 CONCLUSÃO

Os dados do presente trabalho permitem concluir que *SETD4* é uma enzima cuja transcrição encontra-se aumentada no contexto da leucemia linfoblástica aguda e possivelmente também de outras neoplasias, carecendo estas de confirmação por técnicas mais precisas como RT-qPCR. Em LLA, este aumento da transcrição de *SETD4* parece estar ligado a quantidade de células leucêmicas na medula óssea dos pacientes, e tem impacto negativo no prognóstico dos pacientes classificados na categoria de alta transcrição de *SETD4*. Além disso, a expressão de *SETD4* está atrelada à resposta ao tratamento quimioterápico em pacientes com LLA, o que evidencia a importância deste gene como um marcador prognóstico importante com grande potencial de utilidade clínica. De forma semelhante, em leucemia mieloide aguda, câncer de pulmão e carcinoma de estômago, foi observado que a amplificação e o ganho do número de cópias de *SETD4* também têm impacto negativo na sobrevivência. Sabendo que não poucas mutações de médio impacto atingem o domínio SET, não seria uma surpresa que a função da proteína ficasse prejudicada nestes casos, de maneira que estudos futuros podem avaliar essa possibilidade através de análises de perda ou ganho de função. O desenvolvimento de drogas e tratamentos específicos para *SETD4* representam um esforço que pode vir a ser recompensado com a descoberta de uma nova abordagem para o tratamento da LLA, bem como de outras neoplasias. Ademais, a proteína tem sua transcrição altamente correlacionada com a de duas outras metiltransferases: *SMYD2* e *SETMAR*, de maneira que *SETD4* pode compartilhar com a primeira a regulação da transição da fase G1 para a fase S do ciclo celular, e talvez até compartilhar substratos em vias comuns, influenciando os efeitos decorrentes da sua ação. Estudos futuros podem ajudar a desvendar a importância destas correlações, bem como lançar luz sobre os impactos das alterações de *SETD4* na carcinogênese, que descrevemos neste trabalho.

REFERÊNCIAS

1. Leukemia. In: Britannica Academic [Internet]. Encyclopædia Britannica; 2016.
2. Vardiman JW, Thiele J ADEA. The 2008 revision of the WHO classification of myeloid neoplasms and acute leukemia: rationale and important changes. 2015;114(5):937–52.
3. Bhojwani D, Yang JJ, Pui C-H. Biology of Childhood Acute Lymphoblastic Leukemia. *Pediatr Clin North Am*. 2015;62(1):47–60.
4. Pui C-H, Carroll WL, Meshinchi S, Arceci RJ. Biology, risk stratification, and therapy of pediatric acute leukemias: an update. *J Clin Oncol*. 2011;29(5):551–65.
5. Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. *Lancet*. 2013;381(9881):1943–55.
6. Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: A report from the children's oncology group. *J Clin Oncol*. 2012;30(14):1663–9.
7. Hunger SP, Loh ML, Whitlock JA, Winick NJ, Carroll WL, Devidas M, et al. Children's Oncology Group's 2013 blueprint for research: acute lymphoblastic leukemia. *Pediatr Blood Cancer*. 2013 Jun;60(6):957–63.
8. Frishman-levy L, Shemesh A, Bar-sinai A, Ma C, Ni Z, Frenkel S, et al. Central nervous system acute lymphoblastic leukemia : role of natural killer cells1. Frishman-levy L, Shemesh A, Bar-sinai A, Ma C, Ni Z, Frenkel S, et al. Central nervous system acute lymphoblastic leukemia : role of natural killer cells. 2015;125(22):34. 2015;125(22):3420–32.
9. Stoffel T, Nesbit M, Levitt S. Extramedullary involvement of the testes in childhood leukemia. *Cancer*. 1975;345:1203.
10. Pui C, Mullighan CG, Evans WE, Relling M V. Pediatric acute lymphoblastic leukemia : where are we going and how do we get there?. 2012;120(6):1–3.

11. Pieters R, Schrappe M, De Lorenzo P, Hann I, De Rossi G, Felice M, et al. A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial. *Lancet*. 2007;370(9583):240–50.
12. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. *Blood*. 2003;102(7):2321–33.
13. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C, et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A*. 2002;99(12):8242–7.
14. Aydin C, Cetin Z, Manguoglu AE, Tayfun F, Clark OA, Kupesiz A, et al. Evaluation of ETV6/RUNX1 Fusion and Additional Abnormalities Involving ETV6 and/or RUNX1 Genes Using FISH Technique in Patients with Childhood Acute Lymphoblastic Leukemia. *Indian J Hematol Blood Transfus*. 2016;32(2):154–61.
15. Bateman CM, Colman SM, Chaplin T, Young BD, Eden TO, Bhakta M, et al. Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Leukemia*. 2010;115(17):3553–8.
16. Raetz EA, Carroll WL. Refining prognosis in BCR-ABL1-positive ALL. *Blood* [Internet]. 2014;123(11):1626–7.
17. Carroll AJ, Crist WM, Parmley RT, Roper M, Cooper MD, Finley WH. Pre-B Cell Leukemia Associated With Chromosome Translocation 1;19. *Blood*. 1984;63(3):721–4.
18. Burmeister T, Gökbuget N, Schwartz S, Fischer L, Hubert D, Sindram A, et al. Clinical features and prognostic implications of TCF3-PBX1 and ETV6-RUNX1 in adult acute lymphoblastic leukemia. *Haematologica*. 2010;95(2):241–6.
19. Matsuo Y, Drexler HG. Establishment and characterization of human B cell precursor-leukemia cell lines. *Leuk Res*. 1998;22(7):567–79.
20. Zhang Y, Chen A, Yan XM, Huang G. Disordered epigenetic regulation in MLL-related leukemia. *Int J Hematol*. 2012;96(4):428–37.
21. Albert M, Helin K. Histone methyltransferases in cancer. *Semin Cell Dev Biol*.

- 2010;21(2):209–20.
22. Andersson AK, Ma J, Wang J, Chen X, Gedman AL, Dang J, et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet.* 2015;47(4):330–7.
 23. Williams DL, Tsiatis A, Brodeur GM, Look AT, Melvin SL, Bowman WP, et al. Prognostic importance of chromosome number in 136 untreated children with acute lymphoblastic leukemia. *Blood.* 1982;60(4):864–71.
 24. Roberts KG, Mullighan CG. Genomics in acute lymphoblastic leukaemia: insights and treatment implications. *Nat Rev Clin Oncol.* 2015;12(6):344–57.
 25. Paulsson K, Johansson B. High hyperdiploid childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer.* 2009 Ago;48(8):637–60.
 26. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature.* 2007;446(7137):758–64.
 27. Allan RS, Nutt SL. Deciphering the epigenetic code of T lymphocytes. *Immunol Rev.* 2014;261(1):50–61.
 28. Luger K, Mäder W, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature.* 1997;389(6648):251–60.
 29. Cutter AR, Hayes JJ. A brief review of nucleosome structure. *FEBS Lett.* 2015;589(20):2914–22.
 30. Sadakierska-Chudy A, Filip M. A Comprehensive View of the Epigenetic Landscape. Part II: Histone Post-translational Modification, Nucleosome Level, and Chromatin Regulation by ncRNAs. 2014;
 31. Philp A, Rowland T, Perez-Schindler J, Schenk S. Understanding the acetylome: translating targeted proteomics into meaningful physiology. *Am J Physiol Cell Physiol.* 2014;307(9):C763–73.
 32. Berdasco M, Esteller M. Aberrant Epigenetic Landscape in Cancer: How Cellular Identity Goes Awry. *Dev Cell.* 2010;19(5):698–711.

33. Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer*. 2004;4(2):143–53.
34. Esteller M. Epigenetics in cancer. *N Engl J Med*. 2008 Mar 13;358(11):1148–59.
35. Sadakierska-Chudy A, Kostrzewa RM, Filip M. A Comprehensive View of the Epigenetic Landscape Part I: DNA Methylation, Passive and Active DNA Demethylation Pathways and Histone Variants. *Neurotox Res*. 2014;27(1):84–97.
36. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012 Mai 29;13(7):484–92.
37. Schübeler D. Function and information content of DNA methylation. *Nature*. 2015 Jan 15; 517(7534):321-6.
38. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology*. 2013;38(1):23–38.
39. Rivera RM, Ross JW. Epigenetics in fertilization and preimplantation embryo development. *Prog Biophys Mol Biol*. 2013;113(3):423–32.
40. Macdonald WA, Mann MRW. Epigenetic regulation of genomic imprinting from germ line to preimplantation. *Mol Reprod Dev*. 2014;81(2):126–40.
41. Smith BC, Denu JM. Chemical Mechanisms of Histone Lysine and Arginine Modifications. *Biochim Biophys Acta*. 2009;1789(1):45–57.
42. Allfrey V, Faulkner R, Mirsky A. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Sci United States*. 1964;315(1938):786–94.
43. Cela O, Scrima R, Paziienza V, Merla G, Benegiamo G, Augello B, et al. Clock genes-dependent acetylation of complex I sets rhythmic activity of mitochondrial OxPhos. *Biochim Biophys Acta - Mol Cell Res*. 2016;1863(4):596–606.
44. Ganguly S, Coon SL, Klein DC. Control of melatonin synthesis in the mammalian pineal gland: the critical role of serotonin acetylation. *Cell Tissue Res*. 2002 Jul 24 [cited 2016 Jul 25];309(1):127–37.

45. Sen N. Epigenetic regulation of memory by acetylation and methylation of chromatin: implications in neurological disorders, aging, and addiction. *Neuromolecular Med.* 2015 Jun [cited 2016 Jul 25];17(2):97–110.
46. Zhang C, Zhong JF, Stucky A, Chen X-L, Press MF, Zhang X. Histone acetylation: novel target for the treatment of acute lymphoblastic leukemia. *Clin Epigenetics.* 2011;1–10.
47. Florean C, Schnekenburger M, Grandjenette C, Dicato M, Diederich M. Epigenomics of leukemia: from mechanisms to therapeutic applications. *Epigenomics.* 2011;3(5):581–609.
48. Herz HM, Garruss A, Shilatifard A. SET for life: Biochemical activities and biological functions of SET domain-containing proteins. *Trends Biochem Sci.* 2013;38(12):621–39.
49. Montenegro MF, Sanchez-del-Campo L, Fernandez-Perez MP, Saez-Ayala M, Cabezas-Herrera J, Rodriguez-Lopez JN. Targeting the epigenetic machinery of cancer cells. *Oncogene.* 2015 Jan;34(2):135–43.
50. Jenuwein T, Allis CD. Translating the histone code. *Science.* 2001;293(5532):1074–80.
51. Chen H, Lonardi S, Zheng J. Deciphering histone code of transcriptional regulation in malaria parasites by large-scale data mining. *Comput Biol Chem.* 2014;50:3–10.
52. Britton L-MP, Gonzales-Cope M, Zee BM, Garcia B a. Breaking the histone code with quantitative mass spectrometry. *Expert Rev Proteomics.* 2011;8(5):631–43.
53. Dillon SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* 2005;6(8):227.
54. Ng HH, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y, et al. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.* 2002;16(12):1518–27.
55. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, et al.

- Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol.* 2002;12(12):1052–8.
56. Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell.* 2011;146(6):1016–28.
 57. Hassler MR, Egger G. Epigenomics of cancer - Emerging new concepts. *Biochimie.* 2012;94(11):2219–30.
 58. Hanahan D, Weinberg RA. The Hallmarks of Cancer. *Cell.* 2000;100:57–70.
 59. Hanahan D, Weinberg R a. Hallmarks of cancer: The next generation. *Cell.* 2011;144(5):646–74.
 60. Wang X, Chen C-W, Armstrong SA. The role of DOT1L in the maintenance of leukemia gene expression. *Curr Opin Genet Dev.* 2016;36:68–72.
 61. Hamamoto R, Saloura V, Nakamura Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat Publ Gr.* 2015;15(2):110–24.
 62. Arantes Faria JAQ. SET domain-containing Protein 4 (SETD4) is a Newly Identified Cytosolic and Nuclear Lysine Methyltransferase involved in Breast Cancer Cell Proliferation. *J Cancer Sci Ther.* 2013;05(02):58–65.
 63. Petrossian TC, Clarke SG. Uncovering the human methyltransferasome. *Mol Cell Proteomics.* 2011;10(1):M110.000976.
 64. Li GM, Wang YG, Pan Q, Wang J, Fan JG, Sun C. RNAi screening with shRNAs against histone methylation-related genes reveals determinants of sorafenib sensitivity in hepatocellular carcinoma cells. *Int J Clin Exp Pathol.* 2014;7(3):1085–92.
 65. Sakamoto, L. H. T. Papel de metiltransferases de proteínas no desenvolvimento e prognóstico de leucemias linfóides agudas da infância
Papel de metiltransferases de proteínas no desenvolvimento e prognóstico de leucemias linfóides agudas da infância. 2014;
 66. 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. *Leuk Res.* 2014;38(4):496–502.

67. Bagger FO, Sasivarevic D, Sohi SH, Laursen LG, Pundhir S, S nderby CK, et al. BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D917–24.
68. Chalifa-Caspi V. GeneAnnot: Interfacing GeneCards with high-throughput gene expression compendia. *Brief Bioinform.* 2003;4(4):349–60.
69. 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;2(5):401–4.
70. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal [Internet].* 2013;6(269):p11.
71. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012;483(7391):603–7.
72. Haferlach T, Kohlmann A, Wiczorek L, Basso G, Te Kronnie G, B??n?? MC, et al. Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: Report from the international microarray innovations in leukemia study group. *J Clin Oncol.* 2010;28(15):2529–37.
73. 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.*
74. Kashimoto K, Komatsu S, Ichikawa D, Arita T, Konishi H, Nagata H, et al. Overexpression of TRIM44 contributes to malignant outcome in gastric carcinoma. *Cancer Sci.* 2012;103(11):2021–6.
75. 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 Jan;

76. 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;444(7119):629–32.
77. Saddic LA, West LE, Aslanian A, Yates JR, Rubin SM, Gozani O, et al. Methylation of the retinoblastoma tumor suppressor by SMYD2. *J Biol Chem*. 2010;285(48):37733–40.
78. Wang L, Li N, Zhang H, Luo X, Dai J, Zhou S, et al. Structure of human SMYD2 protein reveals the basis of p53 tumor suppressor methylation. *J Biol Chem*. 2011;286(44):38725–37.
79. Cho H-S, Hayami S, Toyokawa G, Maejima K, Yamane Y, Suzuki T, et al. RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. *Neoplasia* [Internet]. 2012;14(6):476–86.
80. Trievel RC, Flynn EM, Houtz RL, Hurley JH. Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSM1. *Nat Struct Biol* [Internet]. 2003 Jul [cited 2016 Jul 26];10(7):545–52.

ANEXO I

Tabela A1 – Número amostral de cada tecido analisado na plataforma GTEX

Tecido	Número amostral	Tecido	Número amostral	Tecido	Número amostral	Tecido	Número amostral
Sangue	393	Putâmen (cérebro)	97	Cólon (sigmoide)	149	Pele (não exposta ao sol)	250
Músculo esquelético	430	Córtex ant. cingulado (cérebro)	84	Trompas uterinas	6	Pituitária	103
Ventrículo esquerdo (coração)	218	Hipocampo (cérebro)	94	Córtex renal	32	Útero	83
Apêndice atrial (coração)	194	Esôfago (muscular)	247	Bexiga	11	Intestino delgado	88
Fígado	119	Núcleo caudado (cérebro)	117	Gland. Salivar menor	57	Cérvix (endocérvix)	5
Pâncreas	171	Junção gastroesofageal	153	Mama	214	Tireoide	323
Artéria (tibial)	332	Córtex frontal (cérebro)	108	Cólon (transverso)	196	Próstata	106
Medula espinhal	71	Tec. Adiposo subcutâneo	350	Ovário	97	Testículos	172
Artéria (aorta)	224	Gland. adrenal	145	Cerebelo	125	-	-
Tec. adip. Visceral (omento)	227	Hipotálamo (cérebro)	96	Hemisfério cerebelar	105	-	-
Fibroblastos transformados	284	Núcleo acumbens (cérebro)	113	Cérvix (ectocérvix)	6	-	-
Substância negra (cérebro)	63	Pulmão	320	Fígado	104	-	-
Amígdala (cérebro)	72	Córtex (cérebro)	114	Pele (exp. ao sol)	357	-	-
Artéria (coronária)	133	Estômago	193	Nervo (tibial)	304	-	-
Linfócitos transformados (EBV)	118	Esôfago (mucosa)	286	Vagina	96	-	-

ANEXO II

Tabela A2 – Valores de P obtidos pela aplicação do teste de log-rank para tendências às curvas obtidas pelo cBioPortal

Câncer	Valor de P	Câncer	Valor de P
Leucemia mieloide aguda	0.0034**	Adenocarcinoma pulmonar	0.9927
Câncer de bexiga	0.3581	CCE pulmão	0.2367
Carcinoma de bexiga urotelial	0.8351	Cistadenocarcinoma seroso de ovário	0.3563
Adenocarcinoma mamário	0.0739	Adenocarcinoma gástrico	0.0606
Carcinoma esofageal	0.1553	CCE cabeça e pescoço	0.5439