



**UNIVERSIDADE FEDERAL DO CEARÁ  
CENTRO DE CIÊNCIAS  
DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR  
BACHARELADO EM BIOTECNOLOGIA**

**MARIA EDUARDA ARAÚJO ESCÓCIA**

**ANÁLISE *IN SILICO* DE VARIANTES GENÉTICAS NOS GENES *REVI* E *POLI* EM  
PACIENTES PORTADORES DE SÍNDROME MIELODISPLÁSICA**

**FORTALEZA**

**2022**

MARIA EDUARDA ARAÚJO ESCÓCIA

ANÁLISE *IN SILICO* DE VARIANTES GENÉTICAS NOS GENES *REVI* E *POLI* EM  
PACIENTES PORTADORES DE SÍNDROME MIELODISPLÁSICA

Monografia apresentada ao Curso de Biotecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Biotecnologia.

Supervisor Técnico-Científico: Roberta Taiane Germano de Oliveira.

Orientador: Prof. Dr. Ronald Feitosa Pinheiro.

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- E73a Escócia, Maria Eduarda Araújo.  
Análise in silico de variantes genéticas nos genes rev1 e poli em pacientes portadores de síndrome mielodisplásica / Maria Eduarda Araújo Escócia. – 2022.  
111 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências, Curso de Biotecnologia, Fortaleza, 2022.  
Orientação: Prof. Dr. Ronald Feitosa Pinheiro .  
Coorientação: Profa. Ma. Roberta Taiane Germano de Oliveira .
1. Síndrome Mielodisplásica. 2. Tolerância a danos no DNA. 3. Síntese Translesão. 4. Sequenciamento de Nova Geração. I. Título.

CDD 661

---

MARIA EDUARDA ARAÚJO ESCÓCIA

ANÁLISE *IN SILICO* DE VARIANTES GENÉTICAS NOS GENES *REVI* E *POLI* EM  
PACIENTES PORTADORES DE SÍNDROME MIELODISPLÁSICA

Monografia apresentada ao Curso de  
Biotecnologia da Universidade Federal do  
Ceará, como requisito parcial à obtenção do  
título de Bacharel em Biotecnologia.

Aprovada em: \_\_/\_\_/\_\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. Ronald Feitosa Pinheiro (Orientador)  
Universidade Federal do Ceará (UFC)

---

Ma. Roberta Taiane Germano de Oliveira  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Nicholas Costa Barroso Lima  
Universidade Federal do Ceará (UFC)

Dedico esse trabalho à minha vó, Maria Escócia Martins, a mulher mais  
extraordinária que eu já conheci.

## AGRADECIMENTOS

À minha vó, Maria, que me ofereceu da forma mais singela e honesta, desde o momento em que eu nasci e ao longo de toda a minha formação, o maior exemplo do que significa não ser, mas tornar-se uma mulher forte. Obrigada por ter me dado a oportunidade de sentir o que é ser amada de verdade, e para sempre. Eu jamais vou conseguir retribuir todo o amor que a senhora me deu, mas espero poder fazer jus ao seu apoio dedicando essa pequena parte da minha trajetória para você.

Aos meus pais, por terem lutado durante tantos anos, da forma que podiam, para garantir uma educação de qualidade para mim, sempre me impulsionando na direção dos meus próprios caminhos, obrigada por terem fornecido o alicerce sobre o qual eu construí todas as minhas conquistas até hoje.

Aos meus amigos que me acompanharam nessa grande jornada chamada Biotecnologia, Breno, Bruno, Caio, Camila, Gabi, Gabs, Ítalo, Laís, Letícia, Lívia Pontes, Lívia Cardoso, Saulo, obrigada pelo companheirismo, pelos perrengues que passamos juntos e pelas descobertas que fizemos juntos também, tenho certeza que vamos dominar o mundo. Um agradecimento excepcional para o Finger, por todas os almoços veganos que compartilhamos, e pela ajuda com as peripécias da bioinformática, você é uma das pessoas mais gentis e inteligentes que eu conheço, amo você!

À minha amiga Lorena, por acreditar em mim mais do que eu acredito em mim mesma. À minha amiga Ariele, por estar sempre ao meu lado, nas angústias e nas vitórias. Eu aprendi sobre o real valor de cultivar uma amizade verdadeira com vocês.

A todos os meus amigos do Laboratório Citogenômica do Câncer, por terem me recepcionado de braços abertos, e por terem me permitido aprender um pouquinho mais sobre a SMD. Eu me redescobri enquanto aluna, IC e pesquisadora ao lado e com a ajuda de cada um de vocês. Um agradecimento especial para a Chris do meu Greg, Amanda, você é uma mulher incrível!

À Roberta, por ter me auxiliado enormemente no desenvolvimento desse trabalho, por ter acreditado em mim lá no comecinho de 2021, quando eu não sabia nem encontrar uma metáfase. Obrigada pela confiança e pela oportunidade de trabalhar com algo que eu amo tanto, e pelo incentivo a sempre me aperfeiçoar. Você é um grande exemplo pra mim!

À Mayara, por sempre se mostrar solícita a tirar minhas dúvidas quando necessário, pela paciência durante os treinamentos na citogenética, pela compreensão e pela empatia. Te admiro não só enquanto profissional e pesquisadora, mas enquanto mulher.

Ao meu orientador, Dr. Ronald, por me permitir acesso à aprendizagem de alta qualidade em seu laboratório, por me mostrar o valor da pesquisa científica quando ela está devidamente compromissada com o bem-estar dos pacientes.

À minha psicóloga, Gladys, que me auxiliou em tantos momentos difíceis de inquietude, obrigada por me acompanhar nas minhas descobertas e por me ajudar a compreender melhor meus próprios sentimentos.

Ao CNPq, por ter me fornecido as bolsas de Iniciação Científica que eu tive durante a graduação e à CAPES, pelo apoio no fomento à pesquisa.

À Universidade Federal do Ceará, pela oportunidade de ser a primeira mulher da minha família a obter um diploma de ensino superior.

*“Ao fim do dia, podemos aguentar muito mais do que pensamos que podemos”*

– Frida Kahlo

## RESUMO

As Síndromes Mielodisplásicas (SMD) constituem um grupo de doenças clonais que acometem as células precursoras hematopoiéticas e resultam em processo hematopoiético ineficiente e ocorrências clínicas distintas, como citopenias periféricas, displasias de uma ou mais linhagens celulares e progressão para leucemia mieloide aguda. O diagnóstico ocorre mais comumente em idosos acima de 60 anos, com incidência de até 50 casos para 100 mil habitantes/ano. Acredita-se que as alterações cromossômicas e mutações gênicas envolvidas na patogênese da SMD, presentes em até 80% dos casos, sejam consequências de lesões no DNA ocasionadas após a exposição a agentes genotóxicos, quimioterápicos e a toxinas ambientais, além de fatores de risco como tabagismo e obesidade, suscitando assim na instabilidade genômica. Os mecanismos de reparo do DNA atuam nas lesões de fita simples e dupla, contudo, quando estas não são reparadas, o mecanismo de tolerância ao dano é ativado para evitar a paralisação do processo replicativo. Ao empregar DNA polimerases específicas, o mecanismo de Síntese Translesão (TLS) representa uma via importante que permite a passagem através e além de lesões mais robustas no DNA, todavia, o mecanismo é propenso a erros e possui potencial mutagênico. O objetivo do presente estudo foi analisar, através de ferramentas *in silico*, o perfil genômico e a patogenicidade de variantes genéticas identificadas nos genes *REVI* e *POLI*, codificadores das DNA polimerases TLS, em pacientes portadores de Síndrome Mielodisplásica, buscando potenciais biomarcadores para a SMD. Foram incluídos no estudo 14 pacientes com SMD diagnosticados em hospital terciário de referência. A análise citogenética foi realizada por Banda-G e as variantes genéticas foram obtidas por sequenciamento-alvo (*NGS Target Sequencing*), em seguida, diferentes bancos de dados foram utilizados para anotação e classificação das variantes e no total 5 ferramentas de bioinformática distintas foram empregadas para análise de patogenicidade *in silico*. Para o gene *REVI*, foi possível identificar 48 variantes, destas, 17 foram classificadas como patogênicas, além disso, as variantes p.Glu571Asp e p.Glu1184\*, que obtiveram máxima previsão de patogenicidade, foram detectadas em todos os 14 pacientes. Em relação a *REVI*, Os pacientes de alto risco diagnosticados com SMD - EB 1 e SMD - EB 2 e cariótipos com aneuploidias e del (7q) apresentaram maior frequência de mutações com alto potencial patogênico, em comparação a pacientes de baixo risco com SMD SA-DU e SMD SA-DM e cariótipos normais, os quais apresentaram baixa frequência de mutações com baixo potencial patogênico. Já para o gene *POLI*, foi possível identificar 28 variantes, dentre elas, 9 foram classificadas como patogênicas e a variante p. Glu122Asp, classificada como deletéria por todos os algoritmos empregados, foi detectada em 9 pacientes. No que concerne esse gene, pacientes de alto risco diagnosticados em estágio avançado (SMD EB-1 e SMD-EB 2) com cariótipos incluindo aneuploidias e del(7q) apresentaram maior frequência de mutações, totalizando um escore preditivo de patogenicidade muito alto, enquanto os pacientes de baixo risco portadores de SMD em estágio inicial apresentaram baixa frequência de mutações porém com alto potencial patogênico. O presente estudo permitiu traçar associações importantes entre o acúmulo das variantes encontradas nos genes *REVI* e *POLI* e o agravamento dos quadros clínicos em SMD, apontando assim para uma possível conexão entre a falha dos mecanismos TLS, causada por mutações deletérias, e a patogênese e evolução dessa doença.

Palavras-chave: Síndrome Mielodisplásica; Tolerância a danos no DNA; Síntese Translesão; Sequenciamento de Nova Geração.

## ABSTRACT

Myelodysplastic Syndromes (MDS) comprise a group of clonal diseases that affect hematopoietic precursor cells and result in an inefficient hematopoietic process with distinct clinical occurrences, such as peripheral cytopenias, dysplasias of one or more cell lines and progression to acute myeloid leukemia. The diagnosis occurs more commonly in elderly individuals over the age of 60 years, with an incidence of up to 50 cases per 100,000 inhabitants/year. It is believed that the chromosomal alterations and gene mutations involved in the pathogenesis of MDS, present in up to 80% of cases, are consequences of DNA damage caused after exposure to genotoxic agents, chemotherapeutics and environmental toxins, in addition to risk factors such as smoking and obesity, thus giving rise to genomic instability. DNA repair mechanisms act on single- and double-stranded lesions, however, when these can not be repaired, the damage tolerance mechanism is activated to avoid stalling of the replicative process. By employing specific DNA polymerases, the Translesion Synthesis (TLS) mechanism represents an important pathway that allows passage through and beyond bulky DNA lesions, however, this mechanism is error-prone and has mutagenic potential. The aim of the present study was to analyze, using *in silico* tools, the genomic profile and pathogenicity of genetic variants identified in the genes *REVI* and *POLI*, which encode TLS DNA polymerases, in patients with Myelodysplastic Syndrome, in order to seek potential biomarkers for MDS. 14 patients diagnosed with MDS in a tertiary referral hospital were included in the study. Cytogenetic analysis was performed by G-Banding and the genetic variants were obtained by target sequencing (NGS Target Sequencing), different databases were used for annotation and classification of variants, and in total 5 different bioinformatics tools were used for *in silico* analysis of pathogenic potential. For the *REVI* gene, we identified 48 variants, of which 17 were classified as pathogenic, in addition, p.Glu571Asp and p.Glu1184\* variants, which obtained maximum pathogenicity prediction, were detected in all 14 patients. Regarding *REVI*, high-risk patients diagnosed with MDS - EB 1 and MDS - EB 2 that showed karyotypes with aneuploidies and del(7q) had a higher frequency of mutations, with high pathogenic potential, compared to low-risk patients with MDS RS-SLD and MDS-MLD showing normal karyotypes, which had a low frequency of mutations, with low pathogenic potential. As for the *POLI* gene, it was possible to identify 28 variants, among them, 9 were classified as pathogenic and the p.Glu122Asp, classified as deleterious by all algorithms, was detected in 9 patients. Regarding this gene, high-risk patients diagnosed at an advanced stage (SMD EB-1 and MDS-EB 2) that showed karyotypes including aneuploidies and del(7q) had a higher frequency of mutations, totalizing a very high predictive pathogenicity score, on the other hand, low-risk patients with early-stage disease had a low frequency of mutations, but with high pathogenic potential. The present study allowed us to trace important associations between the accumulation of variants found in *REVI* and *POLI* genes and the progression of clinical conditions in MDS, thus pointing to a possible connection between the failure of the TLS mechanisms, caused by deleterious mutations, and the pathogenesis and evolution of this disease.

**Keywords:** Myelodysplastic Syndrome; DNA damage tolerance; Translesion Synthesis; Next Generation Sequencing.

## LISTA DE FIGURAS

Figura 1 – Principais alterações citogenéticas encontradas em SMD e suas respectivas categorias de prognóstico de acordo com o IPSS-R.....	22
Figura 2 – Frequência dos genes cujas mutações possuem maior evidência em SMD, bem como as respectivas funções e vias de expressão implicadas.....	25
Figura 3 – Principais vias de resposta ao dano do DNA e a via de Tolerância ao Dano (DDT), onde o mecanismo operante para lesões volumosas é representado pela Síntese Translesão (TLS).....	38
Figura 4 – Rede de associação entre as principais enzimas envolvidas no mecanismo TLS, obtida através da ferramenta computacional de predição funcional de redes de proteínas STRING.....	39
Figura 5 – Esquema representativo do mecanismo de Síntese Translesão e seu potencial mutagênico.....	45
Figura 6 – Ilustração do processo realizado para obtenção dos cariótipos dos pacientes com SMD.....	49
Figura 7 – Representação ilustrativa da metodologia utilizada nos itens 3.3 e 3.4.....	50
Figura 8 – Representação esquemática da metodologia utilizada nos itens 3.5 e 3.6.....	53
Figura 9 – Caracterização das variantes do gene <i>REVI</i> detectadas nos pacientes com SMD.....	61
Figura 10 – Caracterização das variantes do gene <i>POLI</i> detectadas nos pacientes com SMD.....	75
Figura 11 – Esquematização dos principais resultados obtidos no presente estudo.....	81

## LISTA DE TABELAS

Tabela 1 –	Frequência relatada de lesões genéticas em SMD.....	23
Tabela 2 –	Classificação clínica da SMD de acordo com os critérios estabelecidos pela Organização Mundial da Saúde em 2016.....	27
Tabela 3 –	Variáveis e pontuações para valores de prognóstico segundo o IPSS-R.....	29
Tabela 4 –	Categorias de risco e escores prognósticos de acordo com o IPSS-R.....	29
Tabela 5 –	Variantes genéticas e suas consequências para a proteína resultante.....	31
Tabela 6 –	Caracterização das variantes identificadas no gene <i>REVI</i> referentes às amostras dos pacientes diagnosticados com SMD.....	58
Tabela 7 –	Análise <i>in silico</i> de predição de patogenicidade das variantes do tipo <i>missense</i> referentes ao gene <i>REVI</i> .....	65
Tabela 8 –	Análise <i>in silico</i> das variantes detectadas na região intrônica de <i>REVI</i> .....	68
Tabela 9 –	Variantes identificadas no gene <i>POLI</i> referentes às amostras dos pacientes diagnosticados com SMD.....	72
Tabela 10 –	Análise <i>in silico</i> das variantes do tipo <i>missense</i> detectadas para o gene <i>POLI</i> .....	77
Tabela 11 –	Análise <i>in silico</i> das variantes intrônicas detectadas no gene <i>POLI</i> .....	80
Tabela 12 –	Caracterização das variáveis sociodemográficas e classificação clínica dos pacientes com SMD.....	108
Tabela 13 –	Caracterização descritiva das variáveis clínicas dos pacientes com SMD.....	109
Tabela 14 –	Frequência dos resultados citogenéticos e seu impacto prognóstico em pacientes com SMD ao diagnóstico.....	111
Tabela 15 –	Descrição clínica dos pacientes com SMD incluídos no estudo.....	113

## LISTA DE ABREVIATURAS E SIGLAS

AA	<i>Aminoacids</i>
BER	<i>Base excision repair</i>
CCUS	<i>clonal cytopenia of unknown significance</i>
cDNA	<i>DNA complementar</i>
CHIP	<i>Clonal hematopoiesis of indeterminate potential</i>
COSMIC	<i>Catalogue Of Somatic Mutations In Cancer</i>
CTH	<i>Célula-Tronco Hematopoiética</i>
DDR	<i>DNA damage response</i>
DDT	<i>DNA damage tolerance</i>
FISH	<i>Hibridização in situ por fluorescência</i>
GnomAD	<i>The Genome Aggregation Database</i>
GRCh37/ hg19	<i>Genome Reference Consortium build 37/hg19</i>
Hb	<i>Hemoglobina</i>
HR	<i>Recombinação homóloga</i>
HUWC	<i>Hospital Universitário Walter Cantídio</i>
ICUS	<i>Idiopathic cytopenia of undetermined significance</i>
INDEL	<i>Insertion–deletion mutations</i>
IPSS	<i>Indice prognostic score system</i>
KCl	<i>Cloreto de potássio</i>
LMA	<i>Leucemia Mieloide Aguda</i>
LoF	<i>Loss of function</i>
MMR	<i>Reparo de erros de emparelhamento de bases</i>
NCBI	<i>The National Center for Biotechnology Information</i>
NER	<i>Reparo por excisão de nucleotídeos</i>
NGS	<i>Next Generation Sequencing</i>
NHEJ	<i>Junção por extremidades não-homólogas</i>
OMS	<i>Organização Mundial da Saúde</i>
PCR	<i>Reação em cadeia da polimerase</i>
PCNA	<i>Antígeno nuclear de proliferação celular</i>
POLH	<i>DNA polymerase eta</i>
POLI	<i>DNA polymerase iota</i>
POLK	<i>DNA polymerase kappa</i>
POLL	<i>DNA polymerase lambda</i>

POLN	<i>DNA polymerase nu</i>
POLQ	<i>DNA polymerase theta</i>
POLZ	<i>DNA polymerase zeta</i>
PROVEAN	<i>Protein Variation Effect Analyzer</i>
RAD18	<i>E3 ubiquitin-protein ligase</i>
REV1	<i>DNA directed polymerase</i>
REV3L	<i>DNA directed polymerase zeta catalytic subunit</i>
RPMI	<i>Roswell Park Memorial Institute medium</i>
SEER	<i>Surveillance, epidemiology and end Results</i>
SF3B1	<i>Splicing factor 3b subunit 1</i>
SIFT	<i>Sorting Intolerant From Tolerant</i>
SMD	<i>Síndrome Mielodisplásica</i>
SMD - EB1	<i>Síndrome Mielodisplásica com Excesso de Blastos 1</i>
SMD – EB2	<i>Síndrome Mielodisplásica com Excesso de Blastos 2</i>
SMD – SA- DU	<i>Síndrome Mielodisplásica com Sideroblastos em Anel e displasia em única linhagem</i>
SMD- SA- DM	<i>Síndrome Mielodisplásica com Sideroblastos em Anel e displasia multilinhagem</i>
SNP	<i>Single nucleotide polymorphism</i>
SNV	<i>Single nucleotide variant</i>
TLS	<i>Translesion synthesis</i>
TP53	<i>Tumor protein 53</i>
UBC	<i>Ubiquitina C</i>
UBZ	<i>Domínio de ligação à ubiquitina</i>
WPSS	<i>WHO Classification-Based Prognostic Scoring System</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>17</b>
<b>1.1</b>	<b>Aspectos gerais da Síndrome Mielodisplásica.....</b>	<b>17</b>
<b>1.2</b>	<b>Etiologia.....</b>	<b>18</b>
<b>1.3</b>	<b>Epidemiologia.....</b>	<b>18</b>
<b>1.4</b>	<b>Patogênese.....</b>	<b>19</b>
<b>1.5</b>	<b>Citogenética e alterações moleculares na patogênese da SMD.....</b>	<b>21</b>
<b>1.6</b>	<b>Classificação Clínica.....</b>	<b>25</b>
<b>1.7</b>	<b>Estratificação prognóstica da SMD.....</b>	<b>27</b>
<b>1.8</b>	<b>Sequenciamento-alvo e sua importância no diagnóstico da SMD.....</b>	<b>29</b>
<b>1.9</b>	<b>Desafio da Integração Genômica e a Medicina Personalizada aplicada à SMD.....</b>	<b>32</b>
<b>1.10</b>	<b>Lesões ao DNA.....</b>	<b>34</b>
<b>1.11</b>	<b>Mecanismos de Reparo do DNA.....</b>	<b>35</b>
<b>1.12</b>	<b>Mecanismo de Tolerância a Danos no DNA: Síntese Translesão.....</b>	<b>37</b>
<b>1.12.1</b>	<b><i>O potencial mutagênico do mecanismo TLS.....</i></b>	<b>40</b>
<b>1.13</b>	<b>Família Y: DNA Polimerases TLS <math>\eta</math> (POLH), <math>k</math> (POLK), <math>\iota</math> (POLI) e REV1.....</b>	<b>41</b>
<b>1.13.1</b>	<b><i>Pol <math>\iota</math> (POLI) e REV1.....</i></b>	<b>42</b>
<b>2</b>	<b>OBJETIVOS.....</b>	<b>46</b>
<b>2.1</b>	<b>Objetivos Gerais.....</b>	<b>46</b>
<b>2.2</b>	<b>Objetivos Específicos.....</b>	<b>46</b>
<b>3</b>	<b>PACIENTES, MATERIAIS E MÉTODOS .....</b>	<b>47</b>
<b>3.1</b>	<b>Aspectos éticos da Pesquisa.....</b>	<b>47</b>
<b>3.2</b>	<b>Obtenção das amostras de medula óssea e Cariótipo por bandeamento-G.....</b>	<b>47</b>
<b>3.3</b>	<b>Sequenciamento-alvo dos genes <i>REV1</i> e <i>POLI</i>.....</b>	<b>49</b>
<b>3.4</b>	<b>Alinhamento com genoma de referência e filtragem.....</b>	<b>49</b>
<b>3.5</b>	<b>Anotação, classificação e caracterização das variantes identificadas em <i>REV1</i> e <i>POLI</i>.....</b>	<b>50</b>
<b>3.6</b>	<b>Análise funcional <i>in silico</i> das variantes identificadas em <i>REV1</i> e <i>POLI</i>.....</b>	<b>51</b>
<b>3.7</b>	<b>Variáveis clínicas e sociodemográficas analisadas para os pacientes no estudo.....</b>	<b>53</b>

<b>4</b>	<b>RESULTADOS.....</b>	<b>56</b>
<b>4.1</b>	<b>Variantes obtidas para o gene <i>REVI</i>.....</b>	<b>56</b>
<b>4.1.1</b>	<b><i>Classificação e caracterização das variantes do gene REVI.....</i></b>	<b>56</b>
<b>4.1.2</b>	<b><i>Análise in silico da patogenicidade das variantes identificadas na região exônica do gene REVI.....</i></b>	<b>62</b>
<b>4.1.3</b>	<b><i>Análise in silico da patogenicidade de variantes identificadas na região intrônica do gene REVI... ..</i></b>	<b>67</b>
<b>4.2</b>	<b>Variantes obtidas para o gene <i>POLI</i>.....</b>	<b>71</b>
<b>4.2.1</b>	<b><i>Classificação e caracterização das variantes do gene POLI.....</i></b>	<b>71</b>
<b>4.2.2</b>	<b><i>Análise in silico da patogenicidade variantes identificadas na região exônica do gene POLI.....</i></b>	<b>76</b>
<b>4.2.3</b>	<b><i>Análise in silico da patogenicidade das variantes identificadas na região intrônica do gene POLI.....</i></b>	<b>79</b>
<b>5</b>	<b>DISCUSSÃO.....</b>	<b>82</b>
<b>5.1</b>	<b>Variantes patogênicas de <i>REVI</i> e <i>POLI</i> detectadas em um alto número de pacientes representam biomarcadores em potencial para SMD.....</b>	<b>83</b>
<b>5.2</b>	<b>Acúmulo de mutações patogênicas nos genes <i>REVI</i> e <i>POLI</i> em pacientes em estágio avançado da SMD com cariótipo del (7q).....</b>	<b>84</b>
<b>5.3</b>	<b>Nulissomia do cromossomo Y e baixa frequência de variantes patogênicas nos genes <i>REVI</i> e <i>POLI</i>.....</b>	<b>85</b>
<b>5.4</b>	<b>Pacientes em estágio avançado da SMD e cariótipo alterado com del(5q) apresentam perfil genômico misto para os genes <i>REVI</i> e <i>POLI</i>.....</b>	<b>86</b>
<b>5.5</b>	<b>Pacientes em estágio inicial da SMD e cariótipo normal apresentam menor frequência de variantes para os genes <i>REVI</i> e <i>POLI</i>.....</b>	<b>88</b>
<b>6</b>	<b>CONCLUSÕES.....</b>	<b>93</b>
<b>7</b>	<b>LIMITAÇÕES E PERSPECTIVAS FUTURAS .....</b>	<b>94</b>
<b>8</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>95</b>
	<b>APÊNDICE A: TERMO DE CONSENTIMENTO LIVRE ESCLARECIDO.....</b>	<b>103</b>
	<b>APÊNDICE B: CARACTERIZAÇÃO DOS PACIENTES EM RELAÇÃO ÀS VARIÁVEIS CLÍNICAS E SOCIODEMOGRÁFICAS.....</b>	<b>107</b>

# **1 INTRODUÇÃO**

## **1.1 Aspectos gerais da Síndrome Mielodisplásica**

A Síndrome Mielodisplásica (SMD) representa um grupo heterogêneo de disfunções clonais em células-tronco hematopoiéticas (CTH), cuja consequência é um processo hematopoiético ineficiente, caracterizado por diferentes distúrbios de diferenciação e maturação celular, apresentando sinais clínicos distintos como medula óssea hiperclular, citopenias periféricas e níveis variados de displasia em uma ou mais linhagens celulares na medula óssea. Atualmente, o curso de progressão dessa doença é heterogêneo, podendo variar a partir um fenótipo indolente e silencioso ao longo dos anos, que se traduz em sintomas mínimos e citopenias leves, até uma transição mais rápida e de alto risco para Leucemia Mieloide Aguda (LMA), que é observada em 30 a 40% dos casos. Uma característica importante dessa doença é sua predominância em idosos, o que salienta a relevância de um diagnóstico efetivo e precoce nesse contexto de curto tempo hábil para tratamento (ROBIN; DE WITTE, 2019).

Ao afetar múltiplas vias de sinalização, expressão e reparo do DNA, a compreensão da fisiopatologia em SMD ainda persiste como um grande desafio no campo da onco-hematologia, e os pacientes acometidos dividem-se em duas fases distintas: o grupo de baixo-risco, cujo objetivo do tratamento é promover uma melhora do processo hematopoético ineficiente, e o grupo de alto risco, cujo tratamento é implementado a fim de estender o tempo de sobrevida e prevenir a progressão da doença para um quadro leucêmico (MAEDA, 2021).

Tendo em vista esse espectro variável das características clínico-biológicas de neoplasias hematológicas, em 2016 a Organização Mundial da Saúde (OMS) realizou uma revisão da sua classificação para tumores de tecidos hematopoéticos e linfóides, introduzindo refinamentos na avaliação de citopenias e na interpretação de achados morfológicos, clínicos e citogenéticos para SMD, com o objetivo de incluir a crescente e rápida acumulação de informações genéticas no diagnóstico e classificação dessa doença (ARBER et al., 2016).

## **1.2 Etiologia**

Considerada como um distúrbio clonal em células-tronco mieloides, a Síndrome Mielodisplásica pode ocorrer de forma primária (*de novo*) ou secundária, devido a danos causados às células da medula óssea. Diversos fatores etiológicos ambientais, ocupacionais e relacionados ao estilo de vida já foram associados a SMD, à exemplo da exposição a quimioterápicos, em especial a agentes alquilantes, bem como exposição ocupacional a radiação e a toxinas ambientais, como o benzeno, e fatores de risco como tabagismo e obesidade. Destaca-se ainda o papel de fatores genotóxicos como pesticidas e solventes concentrados no excesso de casos reportados em trabalhadores rurais e industriais. A SMD familiar já foi reportada previamente, porém é um acontecimento raro (DOTSON; LEBOWICZ, 2021; COSTA et al., 2021).

Os fatores originários da SMD primária, isto é, a forma da doença que não apresenta causa aparente, ainda não estão totalmente elucidados, mas assume-se que ela ocorra devido a um processo de oncogênese resultante do acúmulo de mutações somáticas. Destaca-se, nesse contexto, a influência do envelhecimento do sistema hematopoiético em pacientes idosos. Já a SMD relacionada ao uso de agentes quimioterápicos, também conhecida como *treatment-related MDS* (t-MDS), é associada a um pior prognóstico em comparação à SMD primária: ocorre de 5 a 7 anos após o uso desses agentes, apresenta achados citogenéticos complexos e comumente progride para Leucemia Mieloide Aguda (LMA). Em um estudo com 112 pacientes portadores de t-MDS, 55% progrediram para LMA, enquanto pacientes com SMD primária apresentaram taxa de progressão para LMA de 30%. A média geral de sobrevivência (*median overall survival*) para SMD secundária ou associada a tratamentos é de apenas 30 semanas (MAGALHÃES, et al., 2018; RAZA; IVERSON, 2014).

### 1.3 Epidemiologia

A SMD é majoritariamente diagnosticada em pacientes idosos, com média de diagnóstico entre 71-79 anos de idade e incidência anual de 4,9 casos a cada 100.000 habitantes, de acordo com dados do programa SEER (Surveillance, Epidemiology and End Results), principal autoridade responsável por dados de incidência de câncer e sobrevivência nos Estados Unidos. Contudo, o número de indivíduos acometidos pela doença aumenta de 20 a 50 casos para 100.000 habitantes anualmente após os 60 anos. Além disso, apresenta-se mais comumente em homens do que em mulheres (razão

sexual 3-4:2), uma razão que também se eleva com o aumento da idade (ROBIN; DE WITTE, 2019; BUCKSTEIN, 2020).

Quanto às manifestações clínicas, a doença normalmente apresenta-se de forma assintomática ou sintomática compreendendo citopenias em uma ou mais linhagens celulares, sendo os sintomas mais comuns fadiga (55% dos casos), febre e infecção (15%) e sangramentos (8%). A classificação em SMD depende fundamentalmente de características clínicas variáveis como morfologia celular, porcentagem de blastos no sangue periférico e na medula óssea, significância e número de citopenias, necessidade transfusional e perfil citogenético, dentre outras. Dessa forma, devido ao seu potencial heterogêneo desafiador no contexto diagnóstico, infere-se que os dados epidemiológicos disponíveis no mundo sejam resultantes de uma sub-representação da sua verdadeira incidência (DOTSON; LEBOWICZ, 2021).

No Brasil, em um estudo epidemiológico realizado a partir de dados extraídos do Sistema Unificado de Saúde (SUS), foram reportadas 4.557 mortes devido à Síndrome Mielodisplásica entre 2014 e 2018. Observou-se que pacientes com idade acima de 80 anos foram os mais afetados (41.2% das mortes), e foi possível identificar maior prevalência de morte em mulheres (51.09% das mortes acima de 80 anos). A taxa de mortalidade também demonstrou-se elevada no grupo entre 70-79 anos, totalizando 28.3% mortes. A menor taxa de mortalidade foi encontrada para o grupo entre 10-14 anos (0,3% das mortes totais) (MARTINS et al., 2021).

Recentemente, uma análise multicêntrica de 1.080 casos de SMD primária, conduzida com habitantes do Brasil (345), Chile (100) e Argentina (635), constituiu o primeiro estudo sul-americano a tentar descrever características clínicas, demográficas e prognósticas em Síndrome Mielodisplásica. Observou-se prevalência dos casos em mulheres na população chilena, cuja média de idade foi menor (64). A população brasileira apresentou marcadores clínicos e epidemiológicos similares à população argentina (média de idade 68 e 69 anos, respectivamente), com uma maior taxa de casos para indivíduos do sexo masculino e predominância do subtipo SMD-DM, de acordo com a revisão da OMS de 2016. A população chilena apresentou prognóstico de maior risco (IPSS e IPSS-R), compatível com as taxas menores de sobrevivência observadas, em comparação às demais populações estudadas (BELLI; PINHEIRO, 2015).

## 1.4 Patogênese

A Síndrome Mielodisplásica compreende um conjunto de distúrbios clonais originados a partir de células-tronco hematopoéticas portadoras de várias mutações genéticas. Em decorrência da vantagem seletiva conferida por tais mutações, ao longo do tempo as células anormais sofrem um processo de expansão clonal na medula óssea, tornando-se mais populosas. As células hematopoéticas resultantes desses clones, por sua vez, possuem mecanismos de proliferação e diferenciação defeituosos, e muitas destas acabam sofrendo apoptose durante a diferenciação. Aliada à diferenciação prejudicada, a hematopoese ineficaz se estabelece como o mecanismo-chave por trás das citopenias periféricas e da desregulação imunológica notavelmente encontradas em pacientes com SMD, ocasionando disfunção dos mecanismos de reparo do DNA, erros na maquinaria do ciclo celular e alteração no processo de apoptose (NAKAJIMA, 2021).

Aproximadamente metade dos pacientes em SMD apresentam anomalias cromossômicas, grande parte afetando o número de cópias genômicas, como deleções, monossomias ou trissomias, ou ainda alterações estruturais como translocações desbalanceadas. Contudo, os demais 50% de pacientes diagnosticados apresentam cariótipo normal, o que aponta para o papel fundamental que mutações somáticas exercem na patogênese da doença. Embora muitas dessas alterações sejam passageiras, isto é, sem consequências diretas para a hematopoese, outras lesões genéticas somáticas denominadas mutações condutoras (*driver mutations*) são as responsáveis por conferir a vantagem seletiva clonal mencionada anteriormente, mecanismo propiciador do aumento da proliferação e resistência à apoptose que viabiliza a expansão clonal de células-tronco hematopoéticas defeituosas, explorando suas propriedades de auto-renovação. O crescimento e disseminação desse clone, por sua vez, tem por fim os mecanismos fisiopatológicos que levam à SMD (BEJAR et al., 2011).

Dessa forma, um modelo de múltiplos passos foi estabelecido para explicar tais mecanismos, o qual afirma que as alterações cromossômicas estruturais e mutações somáticas encontradas nas células tronco hematopoéticas ocorrem primeiramente em um clone anormal e geneticamente instável, que surge em meio a um ambiente medular altamente pró-inflamatório e favorável à tumorigênese, acarretando em distúrbios das ordens genética, citogenética e epigenética. Por conseguinte, genes de vias distintas são

afetados, como codificadores de fatores de splicing de RNA, reguladores epigenéticos envolvidos na metilação de DNA ou modificação de histonas, fatores de transcrição, moléculas de sinalização, genes do complexo coesina e vias responsáveis pela resposta ao dano e reparo de DNA. Dentre esses, encontra-se maior frequência de mutação em fatores de splicing de RNA e reguladores epigenéticos (GARCIA-MANERO, 2015; CHAMSEDDINE et al., 2016).

Do ponto de vista terapêutico, os clones anômalos iniciadores da doença representam um potencial reservatório refratário ao tratamento, capazes de persistir em expansão mesmo após uma resposta clínica inicial à terapia, de forma que a quantidade de anomalias genéticas geralmente aumenta à medida que a doença progride. O aumento da carga desses clones, portanto, está associado a quadros de evolução para Leucemia Mieloide Aguda e maior incidência de recidivas (AWADA et al., 2020).

Apesar do panorama mutacional em SMD ainda não ter sido completamente elucidado, é compreendido que o acúmulo de várias mutações condutoras, pertencentes a diferentes vias celulares, potencialmente induzem o fenótipo da doença, de modo que uma grande parcela dos pacientes apresenta uma combinação complexa de diferentes mutações gênicas, responsáveis pela heterogeneidade clínica observada. Com o advento de técnicas de sequenciamento completo do exoma, foi possível descobrir que até 90% dos pacientes carregam uma mutação somática em pelo menos um gene, enquanto a maioria dos indivíduos diagnosticados é portadora de 2 a 4 mutações (CHIEREGHIN et al., 2021).

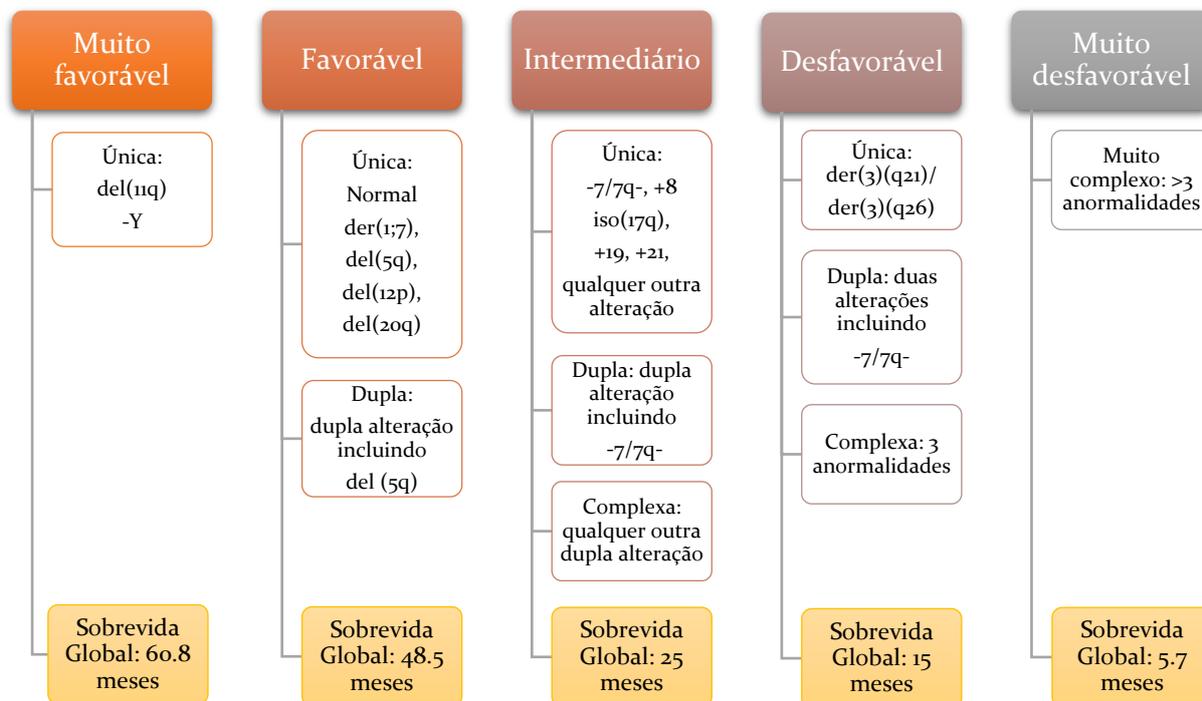
Tal diversidade de mutações somáticas em SMD constitui, por si só, um desafio para extrair interpretações corretas a partir de implicações clínicas significativas. Para além disso, algumas das mutações observadas na SMD sobrepõem-se ao padrão encontrado em indivíduos com hematopoiese clonal que não apresentam citopenias, displasias ou malignidades hematológicas, fenômeno referido como CHIP, assim como em indivíduos com citopenia idiopática de significado indeterminado (ICUS) e indivíduos com citopenia clonal de significado indeterminado (CCUS). Portanto, é extremamente importante distinguir esses fenômenos em indivíduos idosos saudáveis para diferenciá-los da SMD e entender sua relevância clínica para monitoramento posterior (AWADA et al., 2020; BEJAR; GREENBERG, 2017).

## **1.5 Citogenética e alterações moleculares na SMD**

Dentre os diversos testes adicionais necessários para completar a avaliação laboratorial de um paciente portador de SMD, um dos mais fundamentais é a análise citogenética de medula óssea através do exame de cariótipo convencional (bandas-G). Pacientes com SMD primária chegam a apresentar anormalidades citogenéticas em aproximadamente 50% dos casos, e essa taxa pode chegar a 80% dos casos para pacientes com SMD secundária. Tendo em vista a heterogeneidade dos padrões de alterações citogenéticas em SMD, a importância dessa técnica reflete-se tanto no cálculo do prognóstico quanto na decisão da abordagem terapêutica mais efetiva para alguns casos específicos da doença (GARCIA-MANERO, 2020).

A mais recente classificação de risco para SMD é representada por 5 diferentes subgrupos, incluindo alterações citogenéticas distintas, são elas: -7/7q-, 5q-, +8, +19, +21, -Y, -X, i(17q), t(17p), -13/13q-, del(11q), del/t(12p), del(9q), t(11;16), t(3;21), t(1;3), t(2;11), inv/t(3) e t(6 ;9) (Figura 1).

Figura 1 – Principais alterações citogenéticas encontradas em SMD e suas respectivas categorias de prognóstico de acordo com o IPSS-R.



Fonte: Adaptado de SCHANZ et al., 2012.

Além disso, uma parcela significativa dos pacientes já diagnosticados adquire mudanças citogenéticas adicionais ao longo do curso da doença, estando este fenômeno associado ao risco aumentado de progressão para Leucemia Mieloide Aguda e a piores índices de sobrevivência, reafirmando, portanto, a importância do acompanhamento periódico realizado através de exames de cariótipo para identificar alterações cromossômicas potencialmente prejudiciais nesses casos (PERAZZIO; CHAUFFAILLE, 2019).

Outras técnicas podem ser utilizadas para auxiliar no diagnóstico de SMD em amostras com <20 metáfases, como a citometria de fluxo, hibridização *in situ* por fluorescência (FISH) e sequenciamento genético. A citometria de fluxo pode ajudar na identificação de padrões fenotípicos anormais em casos com displasia mínima, contudo, esses casos representam uma baixa porcentagem dos pacientes. Quanto à citogenética molecular, em um estudo realizado por Pinheiro e colaboradores em 2009 foi possível verificar que o uso de FISH acarretou um incremento de apenas 10% de benefício em relação ao cariótipo por bandas-G, revelando assim o baixo custo-benefício dessa técnica e tornando o uso da citogenética convencional o procedimento padrão (PINHEIRO et al., 2019).

Considerando-se o avanço das tecnologias de sequenciamento de nova geração (NGS), tem se ganhado maior conhecimento sobre as mutações comumente associadas à Síndrome Mielodisplásica: alterações em mais de 40 genes foram encontradas nesse contexto, cujas sequências codantes incluem componentes do spliceossomo, fatores de remodelamento da cromatina, moduladores de padrões epigenéticos e distintos fatores de transcrição. Atualmente, pesquisadores são capazes de identificar uma ou mais mutações condutoras em até 80 a 90% dos pacientes, sendo as mais comuns encontradas nos genes *SF3B1*, *TET2*, *SRSF2*, *ASXL1*, *DNMT3A*, *RUNX1*, *U2AF1*, *TP53* e *EZH2* (Tabela 1) (BEJAR et al., 2011).

Tabela 1 - Frequência relatada de lesões genéticas em SMD.

Gene	%	Localização	Função
<i>SF3B1</i>	28	2q33	Fator de splicing
<i>TET2</i>	21	4q24	Controle da hidroximetilação de citosina
<i>ASXL1</i>	14	20q11	Regulador epigenético

<i>SRSF2</i>	12	17q25	Fator de splicing
<i>RUNX1</i>	9	21q22	Fator de transcrição
<i>TP53</i>	8	17p13	Fator de transcrição
<i>U2AF1</i>	7	21q22	Fator de splicing
<i>EZH2</i>	6	7q36	Proteína do grupo Polycomb
<i>NRAS</i>	4	1p13	Transdução de sinal
<i>JAK2</i>	3	9p24	Tirosina quinase
<i>ETV6</i>	3	12p13	Fator de transcrição
<i>CBL</i>	2	11q23	Transdução de sinal
<i>IDH2</i>	2	15q26	Metabolismo celular, regulação epigenética
<i>NPM1</i>	2	5q35	Fosfoproteína
<i>IDH1</i>	1	2q33	Atua como IDH1 (descarboxilação oxidativa do isocitrato)
<i>KRAS</i>	<1	12p12	Transdução de sinal
<i>GNAS</i>	<1	20q13	Proteína-G
<i>PTPN11</i>	<1	12q24	Fosfatase proteica
<i>BRAF</i>	<1	7q34	RAF-quinase
<i>PTEN</i>	<1	10q23	Fosfatase
<i>CDKN2A</i>	<1	9q121	Controle do ciclo celular

Fonte: Adaptado de BEJAR et al., 2011.

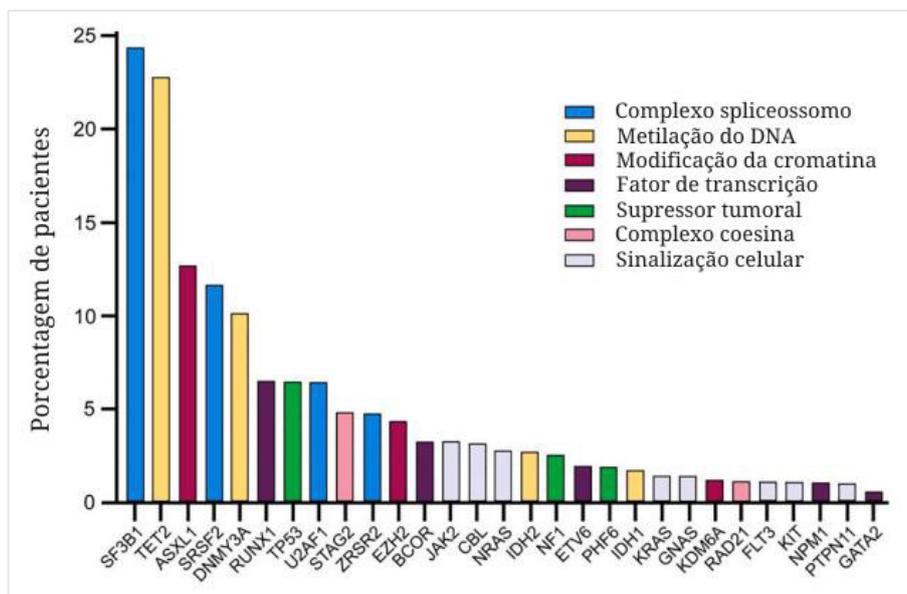
Essas mutações têm sido correlacionadas com diferentes características clínicas, como severidade de citopenias, porcentagem de blastos, achados citogenéticos e sobrevida global. Ressalta-se que mutações genéticas ainda não estão inclusas no sistema de score prognóstico para SMD, todavia, já marcam suas influências na classificação e fisiopatologia da doença: mutações nos genes *RUNX1*, *ASXL1*, *EZH2* e

*SRSF2* por exemplo, são fatores disruptivos da eficiência hematopoética, já o gene *TP53* é um supressor tumoral cuja mutação confere pior prognóstico em comparação às demais mutações, e mutações no gene *SF3B1* identificam um subtipo distinto de SMD caracterizado pela presença de sideroblastos em anel (SMD-RS), que está independentemente associado a um prognóstico favorável, com pacientes alcançando expectativas de vida de mais de 10 anos (BERNARD et al., 2020; CAZZOLA et al., 2013; GREENBERG et al., 2017).

O número de mutações somáticas, bem como a frequência alélica de mutações individuais (VAF) tendem a aumentar à medida que a doença progride de baixo risco para alto risco. Ademais, uma pequena porcentagem (10%) dos casos de SMD estão associados com mutações de linhagens germinativas em vários genes (por exemplo, *GATA2*, *RUNX1*, *DDX41*), denotando assim implicações significativas para a triagem e seleção de doadores em candidatos a transplante de células hematopoéticas (HCT) (KENNEDY; SHIMAMURA, 2019).

No futuro, o uso dessas informações irá certamente redefinir sistemas de escore prognóstico com base em perfis mutacionais únicos e individuais que trarão efeitos contundentes no diagnóstico da doença, uma vez que, nos dias atuais, apesar das consequências funcionais da grande parte dessas mutações ainda não estarem totalmente elucidadas, algumas delas já representam alvos a serem explorados na expansão da compreensão acerca dos mecanismos genéticos por trás da doença (Figura 2), no aperfeiçoamento da classificação prognóstica e na descoberta de novos fármacos que proporcionem terapias mais efetivas e personalizadas (SAYGIN; CARRAWAY, 2021).

Figura 2 – Frequência dos genes cujas mutações possuem maior evidência em SMD, bem como as respectivas funções e vias de expressão implicadas.



Fonte: Adaptado de SAYGIN; CARRAWAY, 2021.

## 1.6 Classificação clínica da SMD

A heterogeneidade de manifestações clínicas previamente mencionada tem sido reconhecida e debatida por profissionais do campo onco-hematológico há muito tempo, e com algumas décadas de esforço os diversos distúrbios clonais que constituem a Síndrome Mielodisplásica foram classificados em diferentes subtipos, com base em características que abrangem achados clínicos, microscópicos e citogenéticos. Em razão da constante atualização de informações acerca dessa doença, os critérios para diagnóstico e classificação da SMD sofreram algumas alterações ao longo dos anos.

Na década de 80, um sistema de classificação houvera sido proposto pelo grupo Franco-Americano Britânico (FAB), no qual uma das principais características distintivas era a proporção de células blásticas no sangue periférico e/ou na medula óssea, sendo portanto um sistema que levava em consideração aspectos puramente morfológicos (BENNETT et al., 1982).

Em 2001, a Organização Mundial da Saúde propôs uma revisão mais robusta à classificação da FAB, com o objetivo de anexar informações atualizadas sobre diferentes manifestações clínicas da doença, achados citogenéticos e aspectos de morfologia celular, citoquímica e imunofenotipagem (VARDIMAN et al., 2002). Tal classificação ainda passaria por uma revisão em 2008, até chegar na versão atual,

proposta em 2016 e vigente até os dias atuais, que adicionou perfis moleculares ao diagnóstico da doença e cujos dados estão dispostos na tabela 2.

Tabela 2 – Classificação clínica da SMD de acordo com os critérios estabelecidos mais recentemente pela Organização Mundial da Saúde em 2016.

Nome	Citopenias e %Blastos no SP	%Sideroblastos em anel	% Blastos MO	Displasias
(SMD-DU)	Uni ou Bicitopenia; $\leq 1\%$ de Blastos	5% de Sideroblastos em anel*	$\leq 5\%$ de blastos	Displasia em 1 linhagem
(SMD-DM)	Citopenia em 1 ou mais linhagens; $\leq 1\%$ de blastos	5% de Sideroblastos em anel*	$\leq 5\%$ de blastos	Displasia em 2 ou 3 linhagens
<b>SMD COM SIDEROBLASTOS EM ANEL (SMD-SA)</b>				
(SMD-SA-DU)	Uni ou Bicitopenia; $\leq 1\%$ de Blastos	$\geq 15\%$ ou $\geq 5\%$ de Sideroblastos em anel*	$\leq 5\%$ de blastos	Displasia em 1 linhagem
(SMD-SA-DM)	Citopenia em 1 ou mais linhagens; $\leq 1\%$ de blastos	$\geq 15\%$ ou $\geq 5\%$ de Sideroblastos em anel*	$\leq 5\%$ de blastos	Displasia em 2 ou 3 linhagens
<b>SMD com Excesso de Blastos (SMD-EB)</b>				
(SMD-EB1)	Citopenia em 1 ou mais linhagens; 2 a 4% de blastos		5 a 9% de blastos	Displasia em 0, 1, 2 ou 3 linhagens
(SMD-EB2)	Citopenia em 1 ou mais linhagens; 5 a 19% de blastos		10 a 19% de blastos	
<b>SMD INCLASSIFICÁVEL (SMD-i)</b>				
SMD-i com 1% de blastos	Citopenia em 1 ou mais linhagens; $\leq 1\%$ <sup>1</sup>		$\leq 5\%$ de blastos	Displasia em 1 ou mais linhagens

	de blastos			
SMD-i com displasia em única linhagem e pancitopenia	Citopenia em 3 linhagens; $\leq 1\%$ de blastos		$\leq 5\%$ de blastos	Displasia em 1 linhagem
SMD-i baseada em alterações Citogenéticas	Citopenia em 1 ou mais linhagens; $\leq 1\%$ de blastos	$\leq 15\%$ de Sideroblastos em anel	$\leq 5\%$ de blastos	Ausência de displasias
<b>SMD COM DEL(5q) ISOLADA</b>	Uni ou Bicitopenia; $\leq 1\%$ de Blastos		$\leq 5\%$ de blastos	Displasia em 1 ou mais linhagens, Del(5q) isolada ou com 1 alteração adicional, exceto del(7q)-7

\*Condição na presença da mutação do gene *SF3B1*; <sup>1</sup>1% de blastos deve ser encontrado em pelo menos duas situações isoladas. SMD-DU: SMD com displasia unilinhagem, SMD-DM: SMD com displasia multilinhagem, SMD-SA-DU: SMD com sideroblastos em anel e displasia unilinhagem, SMD-SA-DM: SMD com sideroblastos em anel e displasia multilinhagem, SMD-EB 1 e 2 (SMD com excesso de blastos 1 e 2).

Fonte: Adaptado de ARBER et al. (2016)

### 1.7 Estratificação prognóstica da SMD

O vocabulário desenvolvido para auxiliar na classificação dos diferentes subtipos de SMD inclui classes de anemias refratárias, anemias com sideroblastos em anel e com excesso de blastos, no entanto, essas denominações por si só não são suficientes para estabelecer decisões quanto ao prognóstico e intervenções terapêuticas adequadas. Tendo em vista que o índice de sobrevida dos pacientes é variável, partindo de apenas alguns meses até anos, uma classificação prognóstica de boa acurácia com base em escores específicos fez-se necessária para SMD, por conseguinte, diferentes sistemas de estratificação prognóstica têm sido utilizados (ROBIN; DE WITTE, 2019)

Uma das ferramentas mais adotadas nesse contexto, o IPSS, foi introduzido por Greenberg e colaboradores em 1997 (GREENBERG et al., 1997). Sua versão revisada, denominada IPSS-R e proposta em 2012 (GREENBERG et al. 2012), tem

sido utilizada como ferramenta padrão para estratificação prognóstica em SMD desde então (Tabela 3). Dividindo-se em 5 categorias de risco: muito baixo, baixo, intermediário, alto e muito alto risco, o sistema atualmente se baseia na proporção de blastos na medula óssea e no número e grau de citopenias, leva em consideração a idade do paciente e incorpora novos escores para achados citogenéticos, objetivando fornecer uma previsão mais próxima da sobrevida esperada para cada paciente e o risco de progressão da doença para LMA (Tabela 4).

Tabela 3 – Variáveis e pontuações para valores de prognóstico segundo o IPSS-R.

IPSS-R: Valores de escore prognóstico							
Variáveis	0	0.5	1	1.5	2	3	4
<b>Prognósticas</b>							
<b>Citogenética</b>	Muito favorável		Favorável		Intermediária	Desfavorável	Muito desfavorável
<b>% Blastos na medula óssea</b>	$\leq 2$		$> 2$ a $< 5$		5 a 10	$> 10$	
<b>Hb (g/dL)</b>	$\geq 10$		8 a $< 10$	$< 8$			
<b>Neutrófilos (mm<sup>3</sup>)</b>	$\geq 800$	$< 800$					
<b>Plaquetas (mm<sup>3</sup>)</b>	$\geq 100$	50 a $< 100$	$< 50$				

Fonte: Adaptado de GREENBERG et al, 2012.

Tabela 4 – Categorias de risco e escores prognósticos de acordo com o IPSS-R.

IPSS-R: Categorias de risco prognóstico			
Categoria de risco	Escore de risco	Sobrevida	Evolução para LMA (Anos)
<b>Muito baixo</b>	$\leq 1,5$	8,8 (7,8 - 9,9)	-
<b>Baixo</b>	$> 1,5 - 3$	5,3 (5 - 5,7)	10,8
<b>Intermediário</b>	$> 3 - 4,5$	3,0 (2,7 - 3,3)	3,2
<b>Alto</b>	$> 4,5 - 6$	1,6 (1,5 - 1,7)	1,4
<b>Muito alto</b>	$> 6$	0,8 (0,7 - 0,8)	0,73

Fonte: Adaptado de GREENBERG et al, 2012.

Apesar do IPSS-R atualmente cumprir papel fundamental na avaliação de risco para SMD, espera-se que um novo sistema de estratificação prognóstica que incorpore dados genômicos seja desenvolvido, dada a confirmação através de recentes estudos dos benefícios de adicionar dados referentes às mutações genéticas no processo de estratificação de risco em comparação aos modelos prognósticos vigentes (BEJAR et al., 2012; KANTARIJAN et al., 2008).

### **1.8 Sequenciamento-alvo e a sua importância no diagnóstico de SMD**

Tais dados genômicos individualizados, os quais tem influenciado diretamente na rotina de manejo clínico e direcionamento de tratamentos na área da oncologia, podem ser obtidos através de diversas técnicas, dentre elas, o sequenciamento de nova geração (NGS) ganha destaque pela sua capacidade de processar uma maior robustez de dados genéticos, o que permite o sequenciamento massivo paralelo de centenas, milhares de genes, ou até mesmo de genomas completos, dentro de um período de tempo relativamente curto, oferecendo assim múltiplas novas oportunidades no campo da medicina personalizada, de modo que as variantes e mutações genéticas detectadas através dessa tecnologia tem sido extensivamente utilizadas no diagnóstico de doenças, bem como no prognóstico, decisões terapêuticas e acompanhamento de pacientes (QIN et al., 2019)

O uso da tecnologia de sequenciamento-alvo (*Targeted Next Generation Sequencing*) é uma das estratégias mais utilizadas para identificar polimorfismos de nucleotídeo único (SNPs) comuns e variantes genéticas raras de interesse clínico, que podem representar fatores de risco para diversos tipos de câncer, além de influenciar na resposta terapêutica a quimioterápicos ou implicar no surgimento de complicações patológicas. Nessa abordagem, sequencia-se um subconjunto de genes específicos associados à patogênese da doença e cuja relevância clínica já se encontra estabelecida, para garantir resultados que sejam implementáveis na rotina diagnóstica em oncologia, de maneira rápida e custo-eficiente (GULILAT et al., 2019).

Para atingir a etapa final de interpretação acerca das mutações detectadas em determinado paciente, empregam-se análises e pipelines de bioinformática para anotação, classificação e caracterização das variantes identificadas a partir do sequenciamento-alvo, que se dividem entre mutações pontuais e indels, distribuídos

entre as regiões codificadoras (exônicas), cujas consequências podem ser praticamente nulas (mutações sinônimas), ou mais graves, como alteração da sequência nucleotídica (variantes *missense*); parada prematura da tradução (variantes *stop gained*); mudança no quadro de leitura (*frameshift*), variantes que comumente resultam em proteínas truncadas, ou encontram-se ainda em regiões não-codificadoras (intrônicas), cujas consequências também podem ser prejudiciais no caso de substituições e deleções que afetam elementos funcionais regulatórios, como acentuadores (*enhancers*) e silenciadores (*silencers*), diretamente envolvidos na regulação do processo de splicing (Tabela 5).

Tabela 5 – Variantes genéticas e suas consequências para a proteína resultante.

<b>Localizadas em regiões codificantes</b>	
Consequência	Definição
<i>Missense</i>	A substituição de um par de bases de DNA altera o códon para outro que codifica aminoácido diferente do aminoácido usual da posição, alterando a sequência primária da proteína.
Sinônima	A substituição de um par de bases de DNA altera o códon para outro que codifica o mesmo aminoácido e não altera a sequência primária da proteína.
<i>Stop gained</i>	A substituição de um par de bases de DNA resulta num códon de parada prematura, sinalizando o final precoce da tradução.
<i>Frameshift</i>	A adição ou a remoção de bases de DNA altera drasticamente o quadro de leitura de um gene.
<b>Localizadas em regiões não-codificantes</b>	
Consequência	Definição
Variante intrônica	Variantes localizadas na região não codificadora do genoma que ainda não foram classificadas e cujo efeito da alteração genética provocada na função da proteína resultante ainda não é conhecido, ou aparentemente nulo.
<i>Deep intronic variant</i>	Variantes localizadas na região não codificadora do genoma cujo efeito da alteração genética provocada pode comprometer regiões regulatórias e genes de RNA não-codificante.
Variante de sítio de splice	Variantes que ocorrem na extremidade éxon-íntron (sítio aceptor/doador de splice), cujo efeito da alteração genética pode comprometer a maquinaria de splicing e resultar na inclusão de

íntrons ou na exclusão de éxons.

---

Adaptado de VAZ-DRAGO et al., 2017.

Tendo em vista a natureza especificamente heterogênea das Síndromes Mielodisplásicas, as quais se distribuem em diferentes subtipos e apresentam manifestações clínicas distintas, que interagem entre si de maneira complexa resultando em categorias de risco e prognósticos variáveis, a investigação da presença de variantes genéticas deletérias que alteram significativamente a função e estrutura da proteína originada, revela-se como uma importante ferramenta de análise. Nesse sentido, a caracterização de mutações somáticas recorrentes via sequenciamento-alvo em SMD auxilia na identificação de variantes patogênicas, sendo particularmente relevante na prática clínica. Inúmeros estudos sugerem que a determinação do status mutacional de múltiplos alvos genéticos resulta em alto poder preditivo prognóstico em SMD, indicando que o uso dessa técnica constitui uma sólida ferramenta de linha de frente para obter um diagnóstico com custo-benefício satisfatório (XU et al., 2017).

O caso mais notório de aplicação de técnicas de sequenciamento-alvo no auxílio do manejo clínico em SMD é o gene *TP53* – um supressor tumoral de destaque que se encontra mais frequentemente mutado em diversos tipo de câncer – cujas consequências das mutações deletérias para pacientes com SMD se traduzem em quadros de alto risco, com rápida progressão para LMA, além da resistência às terapias convencionais e graves desfechos clínicos. Em um estudo realizado por Berbard e colaboradores (2020), onde se buscou determinar as implicações clínicas de mutações no gene *TP53* em 3.324 pacientes com SMD, foi possível correlacionar mutações bialélicas (*multi-hit biallelic targeting*) com fenótipos de cariótipo complexo, quadros de alto risco e desfechos graves, de modo que esse padrão de mutações atualmente se constitui como um fator preditivo independente do IPSS-R para risco de morte e progressão à Leucemia Mieloide Aguda, salientando, portanto, a importância da investigação do status mutacional em *TP53* para a precisão diagnóstica e prognóstica em SMD, bem como na resposta ao tratamento (BERNARD et al., 2020)

### **1.9 Desafio da Integração Genômica e a Medicina Personalizada aplicada à SMD**

Apesar dos esforços recentes, a integração de dados moleculares à uma rotina clínica já bem estabelecida, que utiliza sistemas como o IPSS-R para determinar intervenções terapêuticas em SMD, ainda representa um desafio. Contudo, essas informações adicionais provenientes de testes genéticos claramente influenciam decisões importantes no curso de tratamento da doença, à exemplo da elegibilidade de pacientes para testes clínicos de novas terapias-alvo de acordo com mutações específicas encontradas, reforçando seu caráter essencial na construção de um futuro para a abordagem molecular e personalizada da Síndrome Mielodisplásica (STEENSMA, 2017).

A maioria dos pacientes com SMD (aproximadamente 70%) são diagnosticados com a forma de baixo-risco, relacionada a um curso relativamente favorável da doença, e recebem intervenções terapêuticas menos agressivas, incluindo fatores de crescimento hematopoiéticos como a eritropoetina, suporte transfusional ou simples observação. Todavia, uma parcela desse grupo de baixo-risco acaba desenvolvendo uma forma mais agressiva da doença, com menores índices de sobrevida global. A habilidade de reconhecer pacientes em tal subgrupo específico com piores prognósticos pode ter implicações fundamentais para o aperfeiçoamento e modelagem de terapias apropriadas a determinados riscos (SPAULDING et al., 2020).

Diante de tal conjuntura, um modelo de baixo risco (LR-PSS) que leva em consideração dados moleculares individuais foi validado numa coorte independente de 288 pacientes por Bejar e colaboradores em 2012 (BEJAR et al., 2012), onde se demonstrou pela primeira vez como o status de mutações somáticas nos genes *DNMT3A*, *SF3B1*, *SRSF2*, e *U2AF1*, alterações associadas à trombocitopenia, tiveram considerável influência na classificação prognóstica de pacientes que poderiam se beneficiar de uma intervenção terapêutica precoce, modulada a partir de modificações genéticas específicas encontradas, reiterando assim a relevância e urgência de abordagens que abranjam a Medicina Preventiva e a Medicina da Precisão no tratamento da SMD.

Nesse contexto, o mesmo grupo de pesquisadores apresentou resultados preliminares de 439 pacientes com SMD que haviam tido suas amostras sequenciadas para mutações específicas em 18 genes diferentes, e encontrou-se que 5 genes (*TP53*, *ETV6*, *ASXL1*, *EZH2* e *RUNX1*), quando mutados, forneciam valor prognóstico independente do IPSS, sendo todos marcadores adversos cujas mutações somáticas

efetivamente aumentavam o escore de risco do paciente portador em uma categoria (BEJAR et al., 2011). Desde então, diversos estudos prognósticos têm sido reportados e um grande empenho internacional tem ocorrido sob a tutela do Comitê Internacional para Prognóstico Molecular de SMD (MDS IGW-PM) para avaliar o real valor dos testes genéticos na estratificação de risco da doença (SPAULDING et al., 2020).

Representando esses esforços recentes, um estudo com base em abordagens computacionais foi capaz de definir com sucesso correlações genótipo-fenótipo de 2,043 pacientes de SMD. Com o objetivo de combinar análises moleculares a variáveis clínicas relevantes, a plataforma de inteligência artificial escolhida forneceu um modelo estatístico robusto e inovativo, traduzido numa interface da web disponível para profissionais da saúde que utiliza redes bayesianas para gerar previsões de sobrevida personalizadas, aliando fatores de risco já compreendidos no sistema IPSS-R a informações obtidas a partir do mapeamento genético individualizado de mutações somáticas em mais de 40 genes. A inclusão do perfil genético dos pacientes permitiu uma estratificação de risco de maior acurácia, abrangendo diversas variáveis morfológicas e genômicas em um modelo preditivo complexo, com enorme potencial de aplicação prática de caráter translacional (BERSANELLI et al., 2021).

Isso posto, levanta-se o questionamento: os testes genéticos moleculares irão substituir as ferramentas tradicionais de estratificação de risco e interpretação clínico-patológica em SMD, como o IPSS e o IPSS-R? É provável que não. O cenário atual já aponta para um futuro mais exequível, que preza pelo princípio da soma e não da exclusão mutual, no qual será possível aumentar o valor preditivo de ferramentas prognósticas com a incorporação de resultados moleculares individualizados à rotina clínica, promovendo a utilização de tecnologias recentes, como painéis genéticos e pipelines de bioinformática, na análise de mutações pontuais em genes relacionados à Síndrome Mielodisplásica para melhor direcionar terapias que priorizem a longevidade e a qualidade de vida dos pacientes acometidos (LINDBERG et al., 2020).

### **1.10 Lesões no DNA**

A manutenção da estabilidade genômica é um processo contínuo e crucial para a sobrevivência do indivíduo e sua espécie. Ácidos desoxirribonucleicos são quimicamente instáveis em condições fisiológicas, de modo que danos ao DNA podem

surgir a partir de fontes endógenas, como o ambiente intracelular aquoso e rico em espécies reativas de oxigênio. Ademais, o estresse genotóxico também advém de fatores ambientais, à exemplo da luz ultravioleta, radiação ionizante, metais pesados, poluentes do ar e quimioterápicos (LINDAHL, 1993). Estima-se que a frequência de danos espontâneos causados ao DNA seja da ordem de 10.000 – 100.000 eventos por célula por dia (LINDAHL, 1993; DE BONT; VAN LAREBEKE, 2004).

Em contraste com a maioria das macromoléculas biológicas que, quando danificadas, simplesmente são degradadas e substituídas, o genoma nuclear – o modelo finamente projetado para todas as funções celulares – possui mecanismos de reparo energeticamente dispendiosos para corrigir rapidamente os danos causados ao DNA. Tal dedicação sugere que essas lesões representam um tipo de dano particularmente crítico e prejudicial à homeostase celular (YOUSEFZADEH et al., 2021). Quando não reparadas em tempo hábil, as lesões acarretam em mutações genéticas ou anomalias cromossômicas que comprometem a viabilidade e metabolismo celular, levando a destinos prejudiciais à longo prazo como senescência, neurodegeneração e carcinogênese (LANGIE et al., 2015).

Em eucariotos, muitos mecanismos evoluíram com o objetivo de prevenir esses efeitos deletérios e manter a integridade da informação genética. As alterações em bases dos ácidos nucleicos, por exemplo, são normalmente removidas por vias de reparo por excisão, que atuam combatendo a instabilidade genômica proveniente de lesões no DNA. Contudo, em células humanas, além de nem todas essas vias serem livres de erros, há situações em que o dano em si não é reparado, mas sim contornado por meio de mecanismos de tolerância ou bypass (passagem), executados por DNA polimerases especializadas. Dessa forma, as respostas a tais lesões podem se traduzir em duas maneiras: reparo ou tolerância aos danos (LINDAHL et al., 1999).

### **1.11 Mecanismos de Reparo do DNA**

Para reparar adequadamente os danos ao DNA que continuamente ameaçam a integridade genética e promover uma manutenção fiel dos processos celulares essenciais para o crescimento, a sobrevivência e a hereditariedade biológica, como a replicação do DNA e a divisão celular, as células desenvolveram múltiplos mecanismos, coletivamente referidos como *DNA Damage Response* (DDR) (Figura 3).

Resultantes de uma cascata de sinalização, essa rede é responsável pela detecção, sinalização e reparo dos danos no DNA, de modo que defeitos em genes constituintes da DDR comumente estão associados à carcinogênese e crescimento tumoral. (CHABANON et al., 2021)

No intuito de assegurar que tais obstruções não resultem na instabilidade genômica, a célula exerce mecanismos que, por vezes, sobrepõem-se agindo no reparo e na tolerância a estresses genotóxicos, sendo eles a via de reparo por excisão de nucleotídeos (NER), reparo por excisão de bases (BER) e o reparo de erros de emparelhamento de bases (MMR), dedicados ao reparo de DNA fita simples; e a recombinação homóloga (HR) e junção por extremidades não homólogas (NHEJ), que atuam no reparo de fita dupla de DNA (IYAMA; WILSON, 2013; ASHOUR et al., 2021).

O reparo de fita simples por excisão consiste basicamente na ação de enzimas nucleases especializadas que irão remover a região ou base alterada para, em seguida, DNA polimerases atuarem na síntese de DNA que preencherá a lacuna deixada utilizando a fita não-danificada como molde. A lesão mais comum envolvida nessa via de reparo são os dímeros timina-timina, originados pela ação comprometedora da luz ultravioleta nos processos de replicação e transcrição do DNA (IYAMA; WILSON, 2013). A via NER lida principalmente com lesões provocadas por agentes endógenos, cujas modificações resultam no colapso da forquilha de replicação, atraso na transcrição e distorção da estrutura em hélice do DNA. Dentre os danos reparados pela via BER, encontram-se particularmente os de natureza frequente ocasionados por agentes endógenos, como oxidação, desaminação e hidrólise espontânea. Já o sistema MMR age seletivamente em fitas recém-sintetizadas de DNA contendo nucleotídeos mal-pareados, no reconhecimento e reparo dos *mismatches* e de loops inserção-deleção que surgem durante a replicação ou recombinação homóloga (ASHOUR et al., 2021).

As lesões em DNA de fita dupla (DSBs) são as formas mais deletérias de dano genotóxico, uma vez que, quando não reparadas, ativam respostas ligadas à morte celular e promovem aberrações cromossômicas como translocações, cujas consequências estão implicadas em defeitos neurológicos e imunológicos, além de fenótipos de envelhecimento prematuro e predisposição ao câncer (DIDERICH et al., 2011). São causadas pela ação endógena, através de espécies reativas de oxigênio

(ROS) e eventos que geram diversidade genética, como a meiose e recombinação em anticorpos, ou pela ação exógena com o uso de quimioterápicos. Suas duas maiores vias de reparo compreendem a recombinação homóloga (HR), a qual demanda uma cromátide-irmã para ser executada, e a junção por extremidades não-homólogas (NHEJ), que opera de forma independente do momento do ciclo celular (células em divisão ou não) (WILSON et al., 2007).

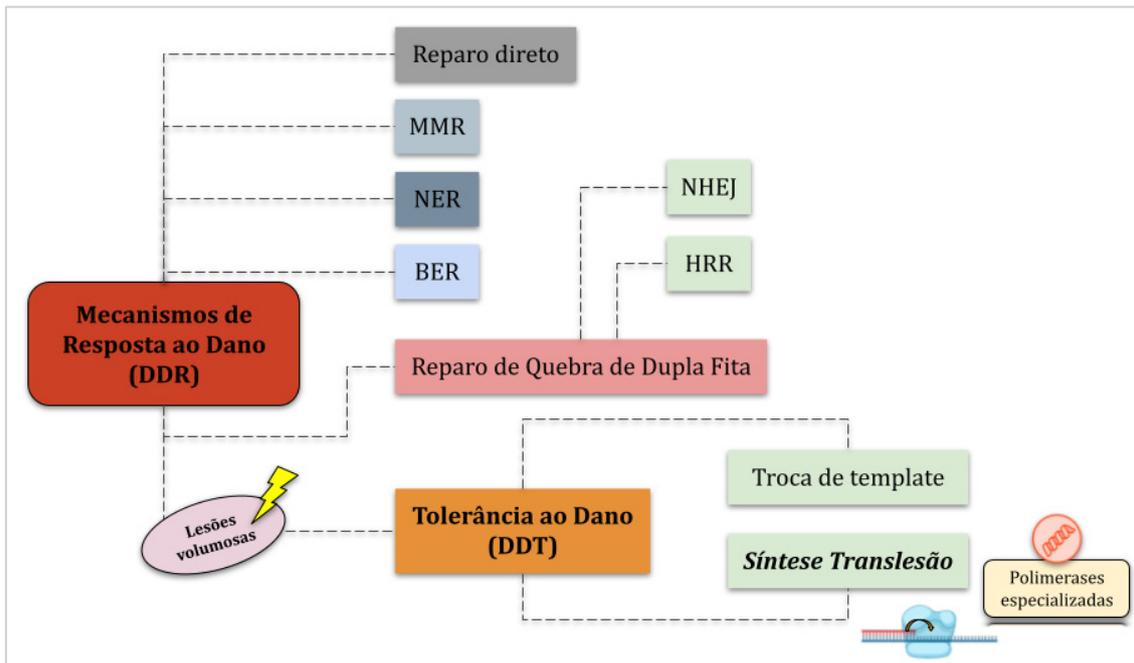
Em 2013, em um estudo abrangendo a genotipagem de 60 pacientes de SMD quanto a genes de reparo do DNA, foi demonstrado pela primeira vez que polimorfismos do gene ATM, o qual atua na sinalização e reconhecimento de lesões de DNA de dupla fita (DSBs), estão associados à patogênese da doença e ao desenvolvimento de SMD de baixo risco (RIBEIRO et al., 2013). Posteriormente, o mesmo grupo de pesquisadores conseguiu associar outros polimorfismos nos genes *RAD51*, *XRCC5* e *XRCC6*, que estão envolvidos nas vias de reparo NHEJ e HR, à patogênese da SMD, ressaltando a relevância desses mecanismos para a manutenção da estabilidade cromossômica em pacientes idosos (RIBEIRO et al., 2015)

### **1.12 Mecanismos de tolerância a danos no DNA: Síntese Translesão**

Após os mecanismos de resposta ao dano do DNA (DDR) serem ativados com o objetivo de reparar as lesões identificadas, as células prosseguem no ciclo celular e seu material genético sofre replicação e posterior transmissão às células-filhas, de forma livre de erros com auxílio das DNA polimerases de alta fidelidade. No entanto, caso essas polimerases clássicas encontrem lesões volumosas na fita molde durante o processo replicativo, a síntese de DNA é interrompida e as forquilhas de replicação são paralisadas. Em decorrência desses eventos, a estagnação prolongada causada é capaz de provocar o colapso das forquilhas de replicação, e, em última instância, ocasionar instabilidade genômica (GHOSAL; CHEN, 2013). Para contornar tal fenômeno e prevenir efeitos deletérios para a célula, mecanismos de tolerância aos danos do DNA (*DNA Damage Tolerance* ou DDT) evoluíram para permitir a replicação genética através e além das lesões encontradas ao longo desse processo, sendo a Síntese Translesão (TLS) uma das suas principais vias de passagem através destes danos em eucariotos (CHANG, 2009). O resumo das principais vias de resposta

ao dano do DNA mencionadas, incluindo os mecanismos DDR e DDT, está disposto na figura 3.

Figura 3 – Principais vias de resposta ao dano do DNA e a via de Tolerância ao Dano (DDT), onde o mecanismo operante para lesões volumosas é representado pela Síntese Translesão (TLS).



Fonte: Adaptado de BLASIAK et al, 2013.

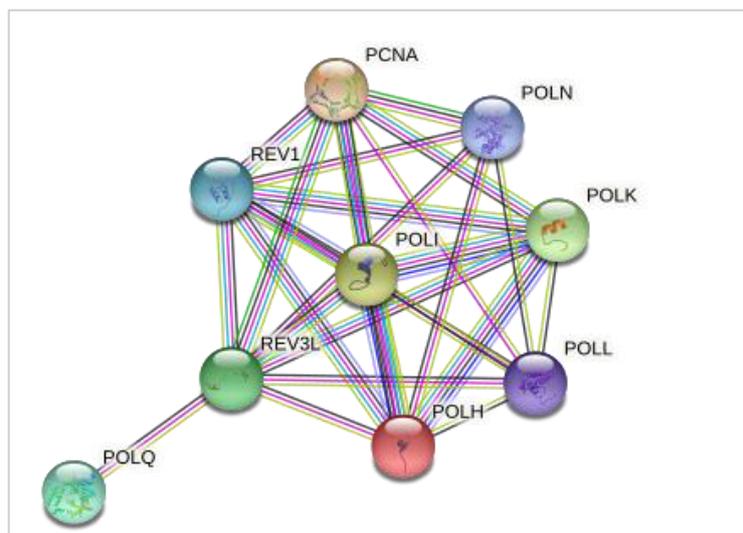
No mecanismo TLS, uma ou mais DNA polimerases não-clássicas são responsáveis pela substituição transitória das polimerases replicativas que estão momentaneamente paralisadas, viabilizando assim a catálise da síntese de DNA através da lesão encontrada. Em geral, essas enzimas possuem maiores sítios catalíticos, que as permitem acomodar lesões volumosas, porém, não possuem domínios de atividade exonucleásica 3' → 5'. Portanto, a via de tolerância ao dano é executada ao custo de uma redução na acurácia replicativa, e, por conseguinte, esse mecanismo de passagem é um fenômeno sujeito a erros e potencialmente mutagênico (SHARMA et al., 2013).

Células de mamíferos possuem pelo menos 7 polimerases TLS de baixa fidelidade: 4 pertencentes à família-Y, Pol η (POLH), Pol ι (POLI), Pol κ (POLK) e REV1, 1 pertencente à família-B, Pol ζ, cuja subunidade catalítica é representada por REV3L, e 2 pertencentes à família-A, Pol θ (POLQ) e Pol ν (POLN) (MAKRIDAKIS

et al., 2012). Cada uma das polimerases TLS possui diferentes especificidades de substrato, atuando em diferentes tipos de danos causados ao DNA. A enzima POLK, por exemplo, mostrou ser eficaz no contorno de lesões mais extensas, como adutos de guanina induzidas por benzopireno, enquanto Pol  $\zeta$  é responsável por transpassar pares de bases distorcidos, como *mismatches* resultantes de inserções errôneas. Dessa forma, a depender da polimerase que será recrutada, as lesões podem ser transpassadas de forma livre de erros, utilizando por exemplo POLH, ou de forma sujeita a erros, como no caso da atividade que envolvem Pol  $\zeta$  e REV1 (QUINET et al., 2018).

Contudo, como a célula “escolhe” qual polimerase será ativada e em quais momentos específicos ainda é um processo não elucidado totalmente. É sabido que tais enzimas interagem entre si e com diversas outras proteínas associadas à replicação celular (Figura 4), reforçando uma hipótese que as polimerases não-clássicas funcionam como um grande complexo multi-proteico, formado após a paralisação da forquilha de replicação, o qual Powers e colaboradores denominaram como complexo de “mutassoma” (replissoma mutagênico) (POWERS et al., 2018).

Figura 4 – Rede de associação entre as principais enzimas envolvidas no mecanismo TLS, obtida através da ferramenta computacional de predição funcional de redes de proteínas STRING.



Legenda: O *cluster* acima exprime uma rede de associação funcional proteína-proteína composta pelas enzimas PCNA, POLN, POLK, POLL, POLH, POLI, POLQ e sua subunidade catalítica REV3L. Os nós da rede são proteínas. As arestas representam as associações funcionais previstas, onde as cores e densidade das linhas indicam diferentes tipos de evidências usadas para prever as associações.

Fonte: Elaborada pela autora, disponível em <https://string-db.org/>.

A monoubiquitinação do antígeno nuclear de proliferação celular (PCNA) parece ser um evento importante na regulação do recrutamento das polimerases TLS pela via de tolerância ao dano. O referente antígeno envolve o DNA de fita dupla e age como uma pinça deslizante, interagindo com grande variedade de proteínas envolvidas na replicação e reparo do DNA. Estudos sugerem que a monoubiquitinação de PCNA promove o contorno direto de lesões pelo recrutamento das polimerases TLS em direção às forquilhas de replicação paralisadas, enquanto a poliubiquitinação do antígeno atua no processo de prevenção ao dano de maneira pouco clara (DAIGAKU et al., 2010; BOEHM et al., 2016).

Apesar da atividade dessas enzimas ser inerentemente sujeita a erros, elas podem mediar reparos livre de erros para lesões específicas, à exemplo de POLH, que consegue ultrapassar lesões induzidas por radiação ultravioleta, de modo que mutações genéticas identificadas no seu gene correspondente resultam em predisposição genética ao câncer de pele (MASUTANI et al., 1999). De forma similar, a enzima POLQ tem um papel preventivo contra o câncer de pele, atuando na progressão da forquilha de replicação em resposta a lesões causadas por luz UV, com atividade sujeita a erros, impedindo a formação de quebras de dupla fita e rearranjos cromossômicos deletérios. Isso posto, infere-se que ambos tipos de polimerases presentes na síntese translesão (sujeita a erros e livre de erros) representam mecanismos preventivos contra a instabilidade genômica (POWERS et al., 2018).

### ***1.12.1 O potencial mutagênico do mecanismo TLS***

Como citado anteriormente, o custo da execução das vias de tolerância ao dano no DNA por vezes se dá através de uma redução na fidelidade replicativa das enzimas utilizadas, as quais podem agir de maneira propensas a erros. A célula, nesse contexto, arrisca-se a tolerar a incorporação errônea de bases no processo replicativo, objetivando evitar a paralisação da forquilha de replicação para assim prevenir a instabilidade cromossômica gerada pelas quebras de fita dupla. Essa solução, porém, pode ocasionar no surgimento de mutações pontuais induzidas por danos ao DNA, sendo o mecanismo TLS especificamente relevante para a oncogênese, um processo

que demanda duas condições profundamente atreladas à atividade da síntese translesão: o acúmulo de mutações somáticas e replicação eficiente do DNA (MAKRIDAKIS et al., 2012).

A ativação exacerbada das vias TLS gera um aumento da mutagênese, podendo levar uma célula normal a se tornar cancerígena ou ainda tornar células cancerígenas resistentes a agentes quimioterápicos. Em contrapartida, a redução na atividade TLS implica numa maior frequência de paralisação das forquilhas de replicação, estabelecendo um risco crítico para a formação de aberrações cromossômicas e, por conseguinte, para o desenvolvimento do câncer. Tendo isso em vista, Ghosal e colaboradores postularam que o mecanismo de síntese translesão na verdade representa uma “faca de dois gumes”, devido ao seu papel conflitante tanto na prevenção quanto na propiciação da instabilidade genômica (GHOSAL; CHEN, 2013).

Ademais, mutações encontradas nos genes correspondentes às polimerases especializadas TLS, bem como alterações nos seus níveis de expressão são eventos notavelmente associados a doenças genéticas raras e predisposição ao câncer, à exemplo das variantes *Xeroderma pigmentosum* (XPV), desordem que causa alta incidência de câncer de pele, tumores ovarianos, adenocarcinomas pulmonares e adenocarcinomas renais (VAZIRI et al., 2021; BROUGHTON et al., 2002), denotando, assim, a importância da realização de estudos mais aprofundados relativos ao mecanismo TLS e suas implicações diretas e indiretas para indivíduos acometidos pelo câncer.

### **1.13 A família-Y: DNA Polimerases TLS $\eta$ (POLH), $k$ (POLK), $\iota$ (POLI) e REV1**

As enzimas constituintes da família-Y de polimerases TLS compartilham cinco motivos de sequências conservadas, formando uma arquitetura estrutural comum semelhante à silhueta de uma mão. A estrutura resultante é composta por domínios do tipo *fingers*, *palm* e *thumb*, implicados diretamente na atividade enzimática de polimerase. O seu domínio N-terminal possui simetria à direita similar às estruturas das polimerases replicativas clássicas, porém com algumas diferenças elementares: não há domínio com atividade exonucleásica 3'-5' e seu sítio ativo é mais solvente e exposto, propriedades que conjuntamente sujeitam sua atuação a uma maior propensão a erros, e em contrapartida conferem a sua habilidade de acomodar e transpassar diversas lesões

no DNA (ZHOU ET AL., 2001). Já a região C-terminal é mais variável e inclui regiões peptídicas de interação-PCNA (PIP boxes), responsáveis por mediar ligações relativamente fracas à PCNA em comparação às polimerases clássicas. Essas enzimas também são constituídas por domínios de ligação à ubiquitina (UBMs) ou dedos de zinco de ligação à ubiquitina (UBZs), os quais se ligam à PCNA monoubiquitinado na posição lisina 164 através do complexo conjugador de ubiquitina Rad6/Rad18 E2/E3. A combinação das regiões PIP e UBDs/UBZs é o fator central que promove especificidade de interação entre as polimerases TLS e o antígeno PCNA na forquilha de replicação (TONZI; HUANG, 2019).

É interessante notar que essa família de enzimas está presente em todos os reinos da vida e que a maioria dos seus componentes são altamente conservados em eucariotos, de modo que todas as enzimas envolvidas na síntese translesão em leveduras possuem evidentes ortólogos em mamíferos, ainda que estes possuam parálogos adicionais a algumas das polimerases provenientes de leveduras (BOITEUX et al., 2013). Tais indícios evolutivos apontam a essencialidade da sua função para o metabolismo celular e sobrevivência biológica.

### **1.13.1 Pol $\iota$ (*POLI*) e *REVI***

A proteína Pol  $\iota$  é uma polimerase especializada de baixa fidelidade, composta por 715 resíduos de aminoácidos, e seu gene codificador (*POLI*) está localizado no cromossomo 18 (posição 18q21). A enzima correspondente é capaz de transpassar diversas lesões no DNA, como dímeros de pirimidina, dano oxidativo e adutos PAH, contudo, ela não é capaz de estender a replicação após a incorporação inicial, sendo necessário para tal passo a utilização de uma polimerase de extensão como Pol  $\zeta$ , pertencente à família-B (WANG et al., 2007; FRANK et al., 2002).

Uma propriedade intrigante desta enzima, que a distingue das demais polimerases é a sua fidelidade, a qual tende a ser baixa em DNA não danificado, porém muito variável com acurácia dependente da sequência molde. As frequências de erro variam desde  $10^{-4}$  a  $10^{-5}$  para template oposto A e  $10^{-1}$  a  $10^{-2}$  para templates opostos G e C. Em templates T, seu mecanismo de ação curiosamente dá preferência à incorporação de uma base que não segue o modelo de pareamento Watson-Crick (adicionando G ao invés de A), um padrão de mutação pontual altamente mutagênico (MAKRIDAKIS et

al., 2012). Devido a esse padrão incorreto de incorporação observado *in vitro*, inicialmente sugeriu-se que Pol  $\iota$  seria uma boa candidata a uma hipermutase somática, participando da hipermutação de genes variáveis de imunoglobulinas *in vivo*. De fato, mais recentemente, foi descoberto o seu papel fundamental em eventos de hipermutação somática em genes de imunoglobulinas, onde Pol  $\iota$  e Pol  $\zeta$  são responsáveis pela geração de mutações em tandem (VAISMAN; WOODGATE, 2020; MAUL et al., 2016).

Desde a descoberta, há mais de 20 anos, a sua real função celular ainda não foi totalmente explicada, contudo, estudos demonstram que Pol  $\iota$  está relacionada à proteção contra danos oxidativos, de modo que variantes identificadas com redução na sua atividade foram capazes de aumentar a susceptibilidade individual de pacientes submetidos a estresse oxidativo via agentes genotóxicos. Além disso, mutações pontuais no gene *POLI* já foram associadas ao aumento do risco de ambos adenocarcinoma pulmonar e carcinoma de células escamosas, relevando um potencial foco de estudo em oncogênese (YEOM et al., 2020; SAKIYAMA et al., 2005).

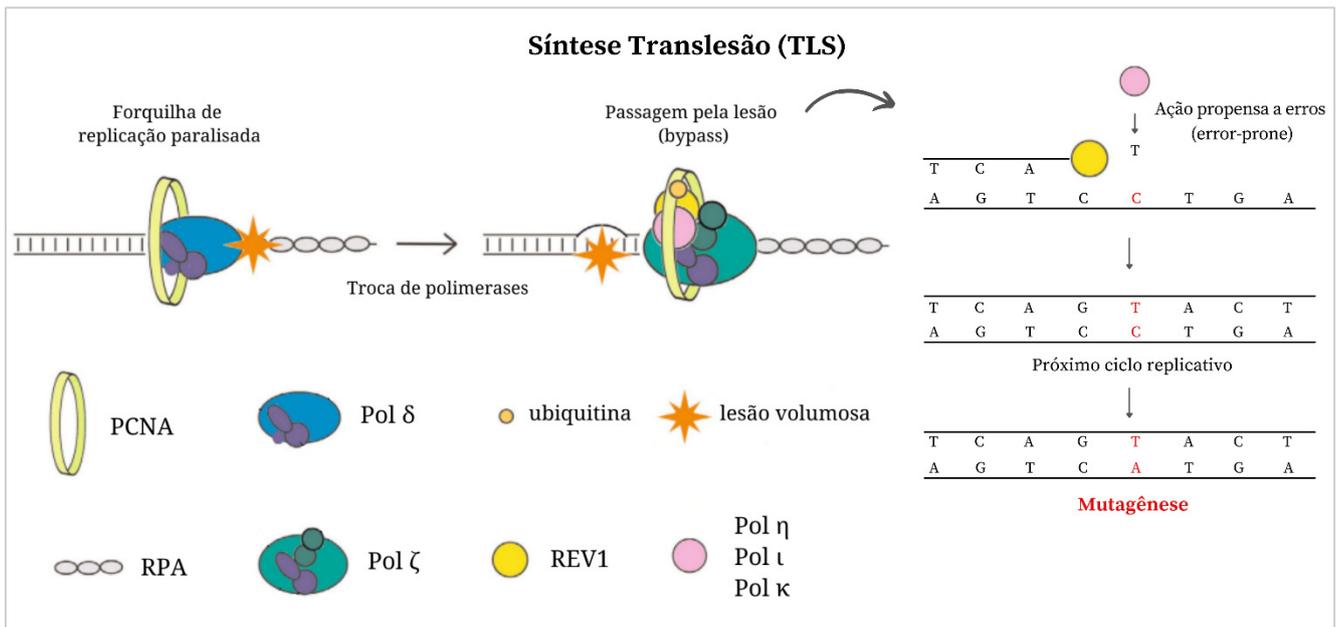
Considerada como o membro mais extremo da família-Y, a proteína REV1 é uma polimerase não clássica monomérica composta por 1251 resíduos de aminoácidos, e seu gene codificador está localizado no cromossomo 2 (2q11.2). Essa enzima possui um domínio polimerase (resíduos 305–738) e duas regiões intrinsecamente desordenadas, uma na porção N-terminal (resíduos 1–304) e outra na porção C-terminal (resíduos 739–985), além de um domínio BRCT, responsável pela ligação à PCNA, em um sítio que se sobrepõe ao motivo de ligação PIP. Sua atividade polimerase incorpora com precisão e eficiência nucleotídeos opostos ao sulco menor, além de adutos de guanina exocíclicos causados pela exposição a vários produtos químicos, como agentes alquilantes, utilizando um mecanismo incomum de síntese de DNA direcionada por modelo, onde um resíduo de arginina do sítio ativo serve como template para incorporação de nucleotídeos. Como REV1 só permite a incorporação eficiente dos nucleotídeos opostos às guaninas danificadas e não danificadas, ela não pode estender a fita para além da lesão molde, assim, desempenha atividade de inserção durante a TLS e requer a assistência de uma polimerase extensora, geralmente Pol  $\zeta$  (Figura 5) (TISSIER et al., 2004; POWERS et al., 2018).

Contudo, REV1 é mais notavelmente reconhecida pela sua atividade regulatória, em oposição à catalítica. Mais especificamente, essa polimerase atua como

uma proteína de *scaffold* para as demais enzimas envolvidas na síntese translesão, tendo uma função importante no agrupamento do complexo multi-proteico formado na TLS, denominado mutassoma. Sua capacidade de interagir simultaneamente com outros componentes do complexo que estão diretamente envolvidos no processo de transpassagem ao dano, como PCNA, Rev7 (constituente da POLZ), Pol  $\eta$  e Pol  $\kappa$  sugere que ela possa facilitar não só o recrutamento, como também a regulação das polimerases TLS, de forma similar à PCNA (TISSIER et al., 2004; JAIN et al., 2018; GUO, 2003).

Ademais, demonstrou-se que a depleção dos níveis de expressão de REV1 em modelos *in vitro* resulta numa menor taxa de indução de mutações com exposição a luz UV, indicando o potencial dessa via como alvo terapêutico para abrandar mutações genéticas (BOUDSOCQ et al., 2002). Outras mutações pontuais identificadas no gene correspondente à REV1 foram associadas a uma pequena proporção de tumores e a um risco aumentado de câncer cervical (HE et al., 2008), o que denota a necessidade de realizar mais investigações acerca do papel de REV1 na etiologia e na progressão do câncer.

Figura 5 – Esquema representativo do mecanismo de Síntese Translesão e seu potencial mutagênico.



Legenda: *Síntese Translesão do DNA. Quando polimerases replicativas clássicas (Pol δ) encontram lesões volumosas no DNA, o resultado frequente é o bloqueio da forquilha de replicação, desencadeando uma cascata de sinalização que culmina na monoubiquitinação de PCNA e no recrutamento de polimerases TLS especializadas, formando assim um complexo de múltiplas subunidades de polimerases e proteínas regulatórias, como RPA. Uma das polimerases da família-Y - iota (Pol ι), eta (Pol η), ou kappa (Pol κ) – incorpora nucleotídeos em oposição ao sítio danificado, às vezes de maneira propensa a erros, enquanto a polimerase zeta (Pol ζ) continua a síntese e extensão da fita de DNA através e sobre a lesão. A principal função de REV1 é coordenar o recrutamento e alocação das polimerases TLS na montagem do complexo multiproteico, interagindo com Pol ζ, polimerases da família-Y e PCNA.*

Fonte: Adaptado de Shilkin et al. (2020)

## 2 OBJETIVOS

### 2.1 Objetivo geral

- Investigar, através de ferramentas *in silico*, o perfil genômico e a patogenicidade de variantes genéticas obtidas a partir de sequenciamento-alvo dos genes *REVI* e *POLI*, codificadores de DNA Polimerases especializadas da via de tolerância ao dano no DNA, em pacientes portadores de Síndrome Mielodisplásica.

### 2.2 Objetivos específicos

- Identificar, caracterizar e classificar as variantes genéticas obtidas por meio de sequenciamento-alvo do gene *REVI* em pacientes com SMD;
- Executar análise *in silico* de predição de patogenicidade das variantes genéticas de *REVI* encontradas nos pacientes com SMD e investigar sua associação à patogênese da doença.
- Identificar, caracterizar e classificar as variantes genéticas obtidas por meio de sequenciamento-alvo do gene *POLI* em pacientes com SMD;
- Executar análise *in silico* de predição de patogenicidade das variantes genéticas de *POLI* encontradas nos pacientes com SMD e investigar sua associação à patogênese da doença.

### **3. PACIENTES, MATERIAIS E MÉTODOS**

#### **3.1 Aspectos éticos da Pesquisa**

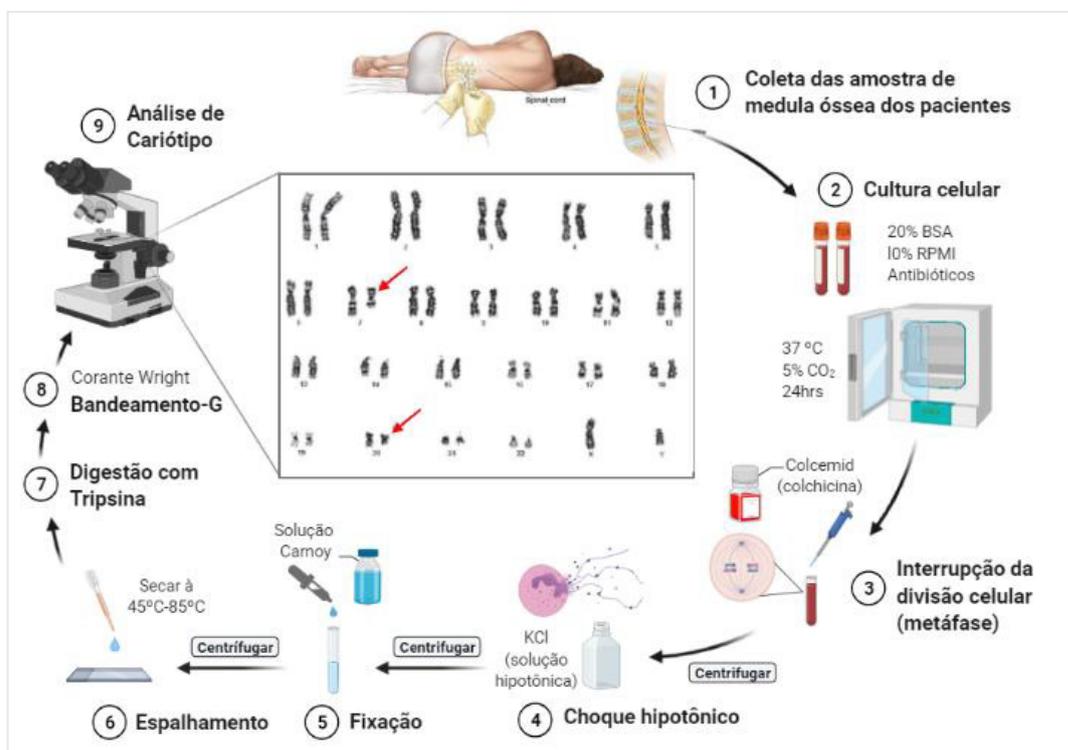
A pesquisa foi submetida e aprovada pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal do Ceará/PROPESQ-UFC (número de processo: 1.292.509) através do sistema da Plataforma Brasil, utilizando-se de Termos de Consentimento Livre e Esclarecido (TCLE) apresentados na sessão apêndice. A equipe executora desta pesquisa comprometeu-se a cumprir todas as diretrizes e normas reguladoras descritas na Resolução n° 466 de 12 de dezembro de 2012 do Conselho Nacional de Saúde que aprova as diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos.

#### **3.2 Obtenção das amostras de medula óssea e Cariótipo por Bandeamento-G**

Os pacientes incluídos neste estudo foram diagnosticados com SMD de acordo com os critérios da Organização Mundial de Saúde de 2016 (ARBER et al., 2016), após terem sido atendidos no Hospital Universitário Walter Cantídio (HUWC) através do Ambulatório de Hematologia da Universidade Federal do Ceará durante o período de 2008 a 2018. Inicialmente, um total de 8 ml de medula óssea foi coletado dos 14 pacientes portadores de SMD em tubos de vidro Vacutainer® contendo EDTA, cujo material foi destinado à extração de RNA e DNA após eles assinarem o termo de consentimento livre e esclarecido (TCLE). Uma segunda amostra de medula óssea é coletada em tubos estéreis contendo heparina, cujo material é destinado à análise citogenética. A cariotipagem dos pacientes foi performada a partir dos protocolos já estabelecidos pelo Laboratório de Citogenômica do Câncer da Universidade Federal do Ceará, segundo a técnica descrita por Chauffaille e adaptada por Pinheiro (PINHEIRO et al., 2009). Após a coleta, as amostras são dispostas em dois frascos contendo 8mL de meio RPMI 1640 (pH 7,0), 2 mL de soro fetal bovino e 100µL de antibióticos (penicilina e estreptomicina). A cultura do material ocorre durante 24 horas em estufa a 37° C. Posteriormente, adiciona-se 40µL de colchicina (Colcemid®) por 30 minutos e depois o material é centrifugado e ressuspenso em solução hipotônica de KCl 0,075M por 20 minutos. Adiciona-se 0,5mL de Carnoy (solução de metanol e ácido acético na

proporção 3:1) e homogeneiza-se a amostra, para centrifugá-la em seguida mais uma vez. Após a centrifugação, lava-se o material 3 vezes com solução de Carnoy (completando-se os volumes de forma seriada 10mL, 8mL e 6mL) sempre centrifugando em seguida para retirar o sobrenadante. Após lavagem, retira-se a solução excedente de cada tubo até um limite proporcional à quantidade do pellet e aplica-se homogeneização para dissolvê-lo. Na etapa de feitura das lâminas, o material presente no tubo é gotejado em lâminas utilizadas para microscopia óptica. Estas, por sua vez, são secadas em placa aquecedora a 45° C até evaporação completa do conteúdo gotejado, depois aquecidas por 10 minutos em placa aquecedora a 85° C. O bandeamento é feito através da técnica de tripsina e as lâminas são coradas com corante Wright. Para concluir a cariotipagem, são analisadas no mínimo 20 metáfases para cada paciente via sistema de captura computadorizada com auxílio do software CytoVision®. Todos os cariótipos medulares obtidos foram confirmados de acordo com o Sistema Internacional de Nomenclatura Citogenética Humana (ISCN) (MCGOWAN-JORDAN, 2016) e a esquematização de todo o processo realizado está disposto na Figura 6.

Figura 6 – Ilustração do processo realizado para obtenção dos cariótipos dos pacientes com SMD.



Fonte: Elaborado pela autora.

### 3.3 Sequenciamento-alvo dos genes *REVI* e *POLI*

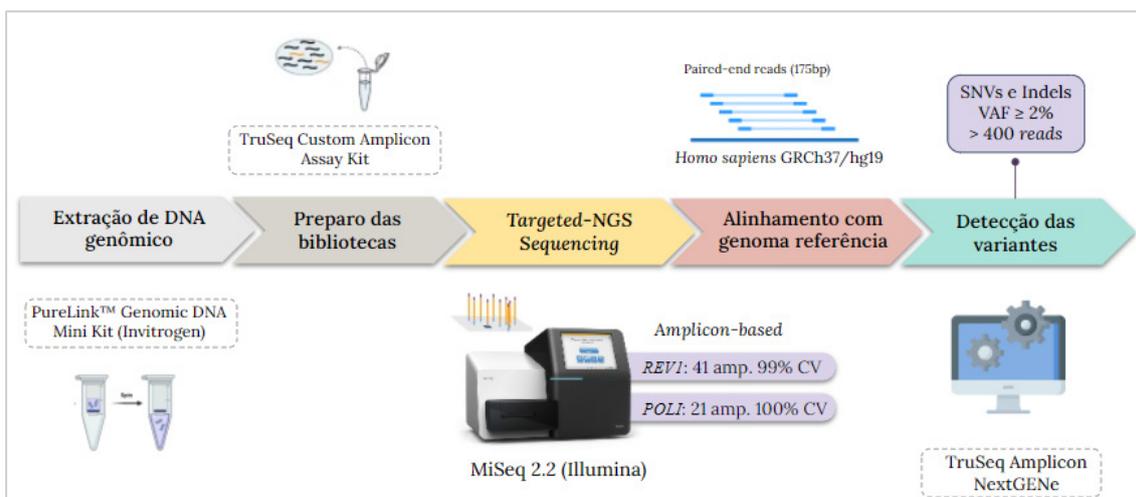
O DNA genômico foi extraído previamente a partir das amostras de medula óssea dos pacientes utilizando-se Kit de Extração de DNA genômico PureLink™ Genomic DNA Mini Kit (Invitrogen) e seguindo as instruções do fabricante.

Em seguida à avaliação da qualidade e concentração das amostras de DNA genômico provenientes dos 14 pacientes diagnosticados com Síndrome Mielodisplásica incluídos neste estudo, o preparo das bibliotecas de DNA foi realizado de acordo com o procedimento padrão recomendado pelo fabricante, Illumina Truseq Custom Amplicon (TSCA), utilizando kit apropriado (TruSeq Custom Amplicon Assay Kit). Após a devida quantificação e normalização das bibliotecas, o sequenciamento-alvo (Targeted-Next Generation Sequencing) dos genes *REVI* e *POLI* foi conduzido no sequenciador MiSeq versão 2.2 (Illumina, San Diego, CA, EUA), procedimento conduzido previamente por Oliveira et al. (2020). Foram sequenciadas as regiões de 41 amplicons do gene *REVI*, com 99% de cobertura, e 21 amplicons do gene *POLI* com 100% de cobertura.

### 3.4 Alinhamento com genoma de referência, filtragem e detecção das variantes

Seguindo a etapa de pré-processamento, as *reads* foram então processadas para remoção de bases de baixa qualidade (Phred quality score <Q30), e o score de qualidade Q30 obtido das bases foi de 93,60%. Os arquivos obtidos no formato FASTQ foram exportados e os fragmentos de DNA paired-end emparelhados (175 bp) foram demultiplexados em arquivos BAM e BAI individuais, alinhados ao genoma referência (*Homo sapiens* UCSC GRCh37/hg19) através do software integrado Illumina Miseq Reporter Software 2.4, utilizando também o software TruSeq Amplicon. Com o auxílio do software NextGENe (<https://softgenetics.com/NextGENe.php>), a etapa de chamada das variantes (*variant calling*) foi realizado, utilizando parâmetros específicos para excluir variantes errôneas, de modo que somente as variações de nucleotídeo único (SNVs) e indels com  $\geq 2\%$  de frequência do alelo variante (VAF) e mais de 400 *reads* de profundidade foram consideradas significativas. A representação ilustrativa do processo metodológico utilizado nos itens 3.3 e 3.4 da referente seção está disposta na Figura 7.

Figura 7 – Representação ilustrativa da metodologia utilizada nos itens 3.3 e 3.4.



Fonte: Elaborada pela autora.

### 3.5 Anotação, classificação e caracterização das variantes identificadas em *REVI* e *POLI*

A fim de comparar as variantes obtidas a partir das amostras de pacientes com SMD presentes no presente estudo com as informações atualmente disponíveis em bases de dados contendo dados sobre SNPs de relevância clínica em oncologia, bem como mutações somáticas e variantes já catalogadas no geral, foram utilizados o banco de dados dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>) e a plataforma COSMIC (*Catalogue of Somatic Mutations in Cancer* - <https://cancer.sanger.ac.uk/cosmic>) para anotação e coleta dos números de identificação (ID - dbSNP e ID - COSMIC) das variantes já conhecidas.

As consequências das variantes de nucleotídeo único (sinônima – substituição de base sem alteração de aminoácido; não-sinônima – substituição de base com alteração de aminoácido, dos tipos *missense* – mudança de aminoácido ou *nonsense* – introdução de códon de parada prematuro), bem como indels (*frameshift* – mudança de quadro de leitura ou *in frame deletion* - sem mudança no quadro de leitura), foram determinadas a partir do software NextGENe, com auxílio de dados presentes nas plataformas dbSNP e Ensembl VEP (*Variant Effect Predictor* - <https://www.ensembl.org/Multi/Tools/VEP>)

A determinação da posição das variantes em relação à sequência de éxons e íntrons de cada gene foi performada utilizando a plataforma NCBI Variation Viewer ([www.ncbi.nlm.nih.gov/variation/view](http://www.ncbi.nlm.nih.gov/variation/view)).

### 3.6 Análise funcional *in silico* das variantes identificadas em *REVI* e *POLI*

Uma análise funcional integrada *in silico* foi conduzida para avaliação da patogenicidade das variantes de nucleotídeo único (Single Nucleotide Variation, SNV) classificadas como não-sinônimas, além das demais variações encontradas como deleções, obtidas através do sequenciamento dos genes *POLI* e *REVI*.

A avaliação acerca da patogenicidade das substituições de aminoácidos provocadas pelas variantes do tipo *missense*, anotadas nas regiões codificadoras dos genes em questão, foi realizada utilizando cinco diferentes algoritmos abertos: SIFT (Sorting Intolerant from Tolerant - <http://sift.bii.a-star.edu.sg/>), o qual prevê se uma substituição de aminoácido afeta a função da proteína com base na homologia de sequência e nas propriedades físicas dos aminoácidos, caracterizando-a como “Prejudicial” ou “Tolerada”; PROVEAN (Protein Variation Effect Analyzer - <http://provean.jcvi.org/index.php>), software que prevê se uma substituição de aminoácido ou *indel* tem impacto na função biológica de uma proteína utilizando abordagem de computação rápida para obter scores de alinhamento de pares de sequências provenientes do genoma anotado na plataforma Ensembl, resulta em variante “Deléteria” ou “Neutra”; Mutation Taster (<http://www.mutationtaster.org/>), que emprega um classificador de Bayes, alimentado constantemente com dados de polimorfismos do *1000 Genomes Project*, na previsão do potencial patogênico de alterações genéticas através do cálculo de probabilidade da mutação ser “Causadora de Doença” ou apenas um “Polimorfismo” inofensivo; MutPred (<http://mutpred.mutdb.org/>), método baseado em *Machine Learning* que integra dados genéticos e moleculares disponíveis nas bases de dados *Human Gene Mutation Database (HGMD)*, SwissVar e dbSNP, desenvolvido para classificar as substituições de aminoácidos como “Patogênicas” ou “Benignas” em humanos, prevendo seu impacto em mais de 50 propriedades diferentes de proteínas.

A partir dessas análises, as mutações do tipo *missense* foram classificadas como patogênicas se dois ou mais softwares fornecessem previsões semelhantes entre

si, adotando-se um score de 0 a 4 de acordo com os resultados obtidos para cada SNV nas ferramentas utilizadas, sendo 4 o valor mais alto e deletério para a previsão de patogenicidade das variantes, de acordo com método descrito por Pousada e colaboradores (POUSADA et al., 2016)

As variantes do tipo mudança de quadro de leitura, *frameshift* e introdução de códon de parada prematuro, *stop gained*, identificadas nas regiões codificantes dos referentes genes foram consideradas como claramente patogênicas, devido às consequências de alto potencial deletério atreladas a essas variantes, as quais comumente comprometem processos transcricionais e traducionais, gerando proteínas truncadas ou não funcionais.

Quanto às variantes de nucleotídeo único e demais alterações encontradas, como deleções, identificadas nas regiões intrônicas dos genes *REVI* e *POLI*, utilizou-se o software baseado em Machine Learning RegulomeDB (<http://www.regulomedb.org>) com o objetivo de interpretar o potencial impacto dessas mutações em funções regulatórias. Tal ferramenta emprega conjuntos de dados públicos curados do GEO (Gene Expression Omnibus), do projeto ENCODE (Encyclopedia of DNA Elements), bem como estratégias de mineração de dados a partir de literatura já publicada e utiliza a anotação de SNPs em elementos regulatórios conhecidos e previstos, em regiões intergênicas do genoma humano que incluem regiões de hipersensibilidade à DNase, sítios de ligação de fatores de transcrição e regiões promotoras que foram bioquimicamente caracterizadas na regulação da transcrição.

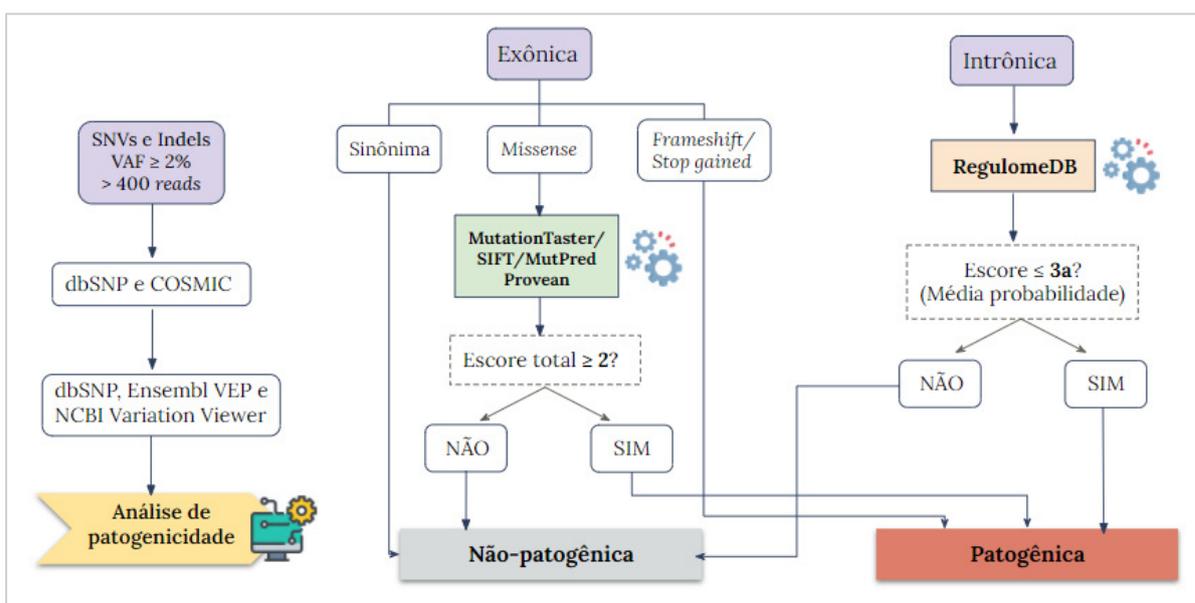
Nesse sentido, RegulomeDB prevê o potencial regulatório de variantes funcionais localizadas em regiões intrônicas estabelecendo um sistema de scores que varia de 1 a 6. Variantes com scores de 1a a 1f indicam probabilidade muito alta de impacto na ligação e expressão de um alvo genético, enquanto scores de 2a a 2c indicam alta probabilidade de afetar a ligação, 3a e 3b indicam média à razoável probabilidade de afetar a ligação e por fim pontuações de 4, 5, 6 indicam evidências mínimas de ligação e 7 nenhum dado disponível. Portanto, para esta análise, quanto mais baixa for a pontuação obtida, maiores são as evidências que a variante de estudo está localizada em uma região funcional, trazendo consequências deletérias para regulação da expressão gênica.

Como ponto de corte para esta análise, foram consideradas como patogênicas as variantes intrônicas identificadas em *REVI* e *POLI* que obtiveram score

a partir de 3a (evidência média de probabilidade de afetar ligação ou regulação gênica) na ferramenta RegulomeDB.

A representação esquemática do processo metodológico utilizado nos itens 3.5 e 3.6 desta seção está disposta na figura 8 a seguir.

Figura 8 – Representação esquemática da metodologia utilizada nos itens 3.5 e 3.6.



Fonte: Elaborada pela autora.

### 3.7 Variáveis clínicas e sociodemográficas analisadas para os pacientes no estudo

A partir das previsões resultantes no que concerne o potencial patogênico das variantes detectadas, os dados obtidos foram então comparados com as variáveis clínicas de cada paciente incluído no estudo descritas abaixo, as quais foram retiradas previamente do banco de pacientes do Laboratório Citogenômica do Câncer, contendo informações sobre seus respectivos prontuários e exames clínicos, buscando investigar a possível relação entre a presença de SNVs e indels com alto potencial patogênico e o risco de progressão dos quadros clínicos consequente agravamento da doença, no intuito de identificar potenciais biomarcadores prognósticos para a SMD.

✓ Sexo: Feminino e Masculino;

- ✓ Idade:  $\leq 60$  anos e  $> 60$  anos;
- ✓ Origem: Urbana e Rural;
- ✓ Histórico de exposição a tóxicos: Com exposição a tóxicos, Sem exposição a tóxicos;
- ✓ Idade categorizada de acordo com IPSS-R:  $\leq 60$  anos,  $> 60 - \leq 70$  anos,  $> 70$  anos -  $\leq 80$  anos e  $> 80$  anos (GREENBERG et al., 2012);
- ✓ Classificação do número de citopenias no sangue periférico: 0 citopenias, 1 citopenia, 2 citopenias e 3 citopenias;
- ✓ Classificação dos valores de hemoglobina (Hb) (IPSS-R):  $\geq 10$ g/dL, 8- 15%;
- ✓ Classificação dos valores de neutrófilos (ANC) (IPSS-R):  $\geq 800$  por  $\text{mm}^3$  e  $< 800$  por  $\text{mm}^3$ ;
- ✓ Classificação dos valores de plaquetas (IPSS-R):  $\geq 100.000$  por  $\text{mm}^3$ , 50.000 -  $< 100.000$  por  $\text{mm}^3$  e  $\leq 50.000$  por  $\text{mm}^3$ ;
- ✓ Celularidade: Hipocelular, Normocelular e Hiper celular;
- ✓ Classificação do número de displasias na medula óssea: 1 displasia, 2 displasias e 3 displasias;
- ✓ Diseritropoiese: Ausência e Presença;
- ✓ Disgranulopoiese: Ausência e Presença;
- ✓ Dismegacariopoiese: Ausência e Presença;
- ✓ Faixas de porcentagem de blastos na medula (IPSS-R):  $\leq 2\%$ ,  $> 2$  a  $< 5\%$ ,  $> 5\%$  -  $\leq 10\%$  e  $> 10\%$ ;
- ✓ Faixa de porcentagem de blastos na medula categorizada em 5%:  $\leq 5\%$  e  $> 5\%$
- ✓ Cariótipo: Normal e Alterado;
- ✓ Cariótipo (Aneuploidia): Alterado com aneuploidia e alterado sem aneuploidia;
- ✓ Cariótipo (Alteração no cromossomo 7): Alterado com del(7q) e alterado sem del(7q);
- ✓ Cariótipo (Alteração no cromossomo 5): Alterado com del(5q) e alterado sem del(5q);
- ✓ Classificação do grupo de risco de acordo com o IPSS-R: Risco muito baixo, baixo risco, intermediário, alto risco e risco muito alto;

- ✓ Quanto a classificação da OMS (2016): SMD-SA-DU (SMD com Sideroblastos em Anel e Displasia de Única linhagem), SMD-SA-DM (SMD com Sideroblastos em Anel e Displasia de Múltipla Linhagem), SMD-EB 1 (SMD com Excesso de Blastos 2) e SMD-EB 2 (SMD com Excesso de Blastos 2);
- ✓ Formas da doença: Iniciais (SMD-SA-DU e SMD-SA-DM) e Avançadas (SMD-EB 1 E SMD-EB 2);
- ✓ Dependência Transfusional: Ausência e Presença [utilizando-se o critério de 1 transfusão a cada 8 semanas durante período de 4 meses (MALCOVATI et al., 2005)];
- ✓ Evolução para LMA: Ausência e Presença.
- ✓ Óbito: Sim e Não.

A caracterização dos 14 pacientes incluídos no presente estudo em relação às variáveis clínicas e sociodemográficas apresentadas encontra-se disposta de forma mais detalhada no Apêndice B.

## 4 RESULTADOS

### 4.1 Variantes obtidas para o gene REV1

#### 4.1.1 Classificação e caracterização das variantes

Após o sequenciamento alvo e análises iniciais de filtragem e caracterização dos dados genômicos extraídos a partir das amostras dos pacientes no presente estudo, foi possível identificar 31 variantes exônicas, e 17 variantes intrônicas para o gene *REV1*, totalizando 48 variantes para esse gene. A distribuição da localização das variantes entre os 23 éxons presentes na sequência codificadora de *REV1* demonstrou uma maior concentração de mutações na região *upstream* do gene, estando a maior frequência de variantes localizada no éxon 6. Enquanto isso, para as variantes intrônicas houve maior concentração de variantes para localizações próximas à região *downstream* do referido gene, cujas variantes foram detectadas mais frequentemente nos íntrons 13 e 18 (Figura 9).

Dentre as variantes exônicas, identificaram-se 29 SNVs, classificados de acordo com sua consequência para a proteína: 21 variantes pontuais *missense* e 7 variantes pontuais sinônimas; além de 1 variante pontual *stop gained* (LoF); e 2 deleções do tipo *frameshift* (mudança no quadro de leitura – *LoF*). De acordo com a anotação das referidas variantes no banco de dados dbSNP, 9 variantes exônicas identificadas nos pacientes com SMD já haviam sido registradas previamente como polimorfismos de nucleotídeo único – SNPs; rs370682495, rs1032393262, rs3087399, rs3087386, rs3087403 e rs3087401 do tipo *missense*; e rs756401282, rs1367690549 e rs2305354, do tipo sinônimas. Em meio às variantes já registradas, 4 delas também já haviam sido reportadas na plataforma COSMIC (COSM4001088, COSM5114887, COSM4001089 e COSM3757517), indicando mutações que possivelmente viraram polimorfismos. Para as demais variantes exônicas encontradas, ainda não havia dados disponíveis nas plataformas mencionadas (Tabela 6).

Quanto às variantes identificadas em regiões não-codificadoras do gene, estas foram classificadas majoritariamente (13) como SNVs do tipo variante intrônica simples, além de 1 variante localizada em sítio doador de splice (variante doador de splice). Ademais, identificaram-se outras 3 variantes do tipo indel, configurando deleções de uma ou mais bases. Dez variantes intrônicas já haviam sido registradas previamente como polimorfismos de nucleotídeo único – SNPs no banco de dados dbSNP, sendo elas rs717454, rs28369944, rs28369943, rs13394927, rs13409359, rs10175852, rs1677058100, rs138292917,

rs1015664521 e rs201548583. Além disso, 3 das variantes intrônicas já constavam na plataforma COSMIC (COSN15304847, COSN6670344, COSN8867434). As informações acerca da classificação, localização, tipo, consequência resultante para a proteína, registros encontrados nos bancos de dados e número de amostras nas quais as variantes mencionadas foram detectadas estão dispostas mais detalhadamente na Tabela 6.

Tabela 6 – Caracterização das variantes identificadas no gene *REVI* referentes às amostras dos pacientes diagnosticados com SMD.

Gene	Tipo de variante	Posição genômica (GRCh37/hg19)	Al. Ref.	Al. Alt.	VAF	Localização	Consequência	Mudança de AA	dbSNP ID	COSMIC ID	N Amostras detectadas
<i>REVI</i>	SNV	2:100038079	T	G	6,80%	Exon 11	<i>Missense</i>	p.Glu571Asp	-	-	14
<i>REVI</i>	SNV	2:100018801	C	A	29,20%	Exon 22	<i>LoF - (Stop gained)</i>	p.Glu1184*	-	-	14
<i>REVI</i>	SNV	2:100055721	G	T	20,70%	Exon 6	Sinônima	p.Gly185=	-	-	14
<i>REVI</i>	SNP	2:100019496	A	G	56,70%	Exon 20	Sinônima	p.Ile1080=	rs2305354	COSM4001088	10
<i>REVI</i>	SNP	2:100081323	C	T	54,50%	Intron 2	Variante intron	p.?	rs28369943	COSN8867434	10
<i>REVI</i>	SNV	2:100027445	A	C	5,80%	Intron 13	Variante intron	p.?	-	-	10
<i>REVI</i>	SNP	2:100022772	T	C	41,90%	Intron 16	Variante intron	p.?	rs717454	COSN15304847	9
<i>REVI</i>	SNP	2:100055506	A	G	50,30%	Exon 6	<i>Missense</i>	p.Phe257Ser	rs3087386	COSM4001089	9
<i>REVI</i>	SNV	2:100081317	T	C	41,90%	Intron 2	Variante intron	p.?	rs28369944	COSN6670344	3
<i>REVI</i>	SNP	2:100022617	G	A	20,40%	Intron 16	Variante intron	p.?	rs13394927	-	3
<i>REVI</i>	SNP	2:100058870	C	T	46,20%	Exon 5	<i>Missense</i>	p.Val138Met	rs3087403	COSM3757517	3
<i>REVI</i>	SNP	2:100055158	T	C	16,20%	Exon 6	<i>Missense</i>	p.Asn373Ser	rs3087399	-	2
<i>REVI</i>	SNV	2:100022422	T	G	5,10%	Exon 17	<i>Missense</i>	p.Ser921Arg	-	-	2
<i>REVI</i>	SNP	2:100055033	C	A	10,30%	Intron 6	Variante intron	p.?	rs10175852	-	2
<i>REVI</i>	SNP	2:100050740	A	C	30,10%	Intron 8	Variante intron	p.?	rs13409359	-	2
<i>REVI</i>	SNV	2:100019220	C	A	25%	Exon 21	<i>Missense</i>	p.Ser1143Ile	-	-	2
<i>REVI</i>	SNV	2:100037963	A	T	5,20%	Exon 11	<i>Missense</i>	p.Ile610Asn	-	-	1
<i>REVI</i>	SNV	2:100029218	A	T	9,70%	Exon 13	<i>Missense</i>	p.Ile716Asn	-	-	1
<i>REVI</i>	SNV	2:100017773	A	T	30,20%	Exon 23	<i>Missense</i>	p.Phe1229Leu	-	-	1

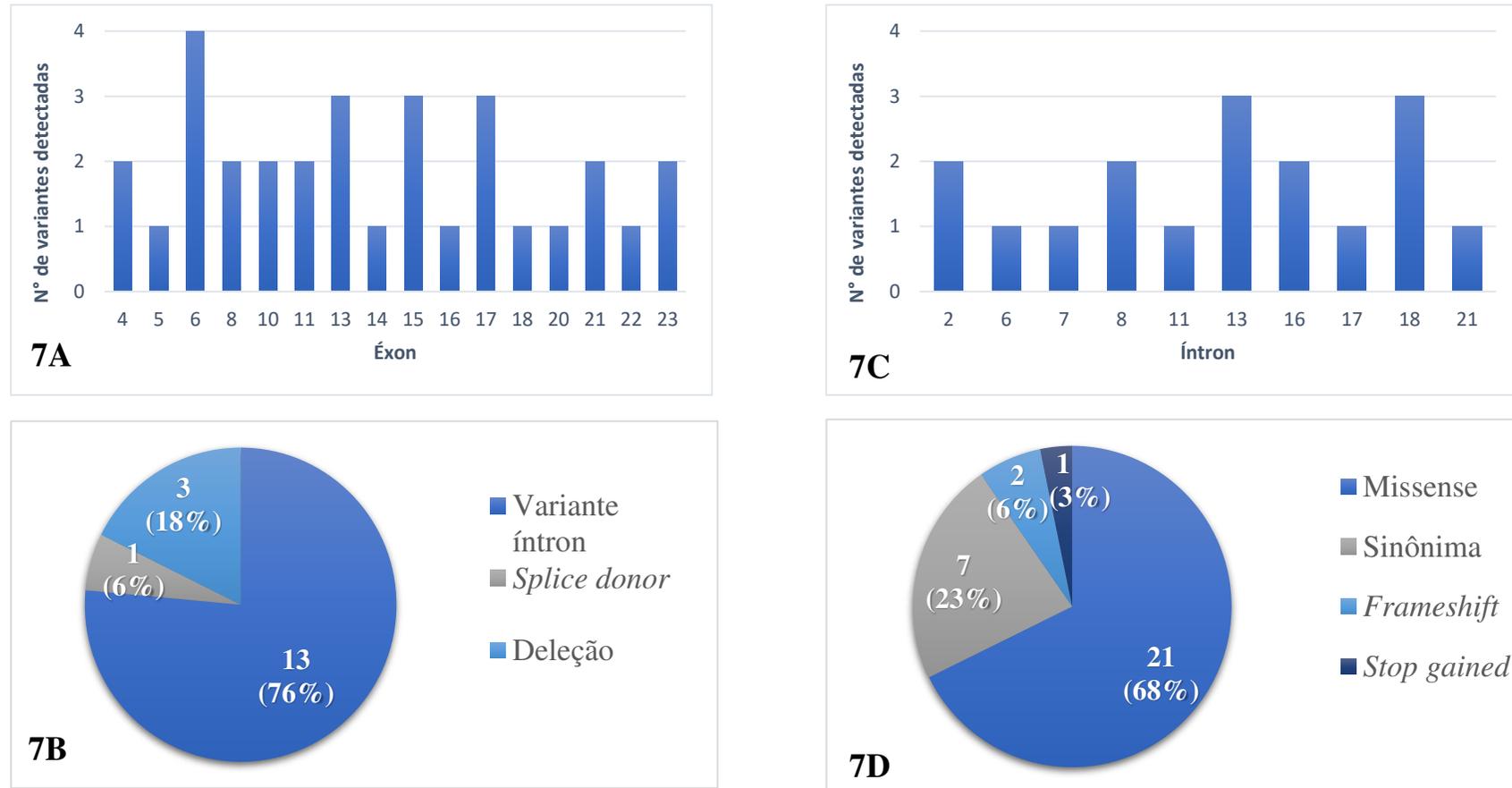
REVI	SNV	2:100065874	C	T	5,40%	Exon 4	Missense	p.Arg92Thr	-	-	1
REVI	SNP	2:100065932	C	T	51,40%	Exon 4	Missense	p.Met72Ile	rs370682495	-	1
REVI	SNP	2:100055171	T	G	50,90%	Exon 6	Missense	p.Thr369Pro	rs1032393262	-	1
REVI	SNV	2:100027255	C	A	9,90%	Exon 14	Missense	p.Lys749Asn	-	-	1
REVI	SNV	2:100017755	A	T	23,90%	Exon 23	Missense	p.Asn1235Lys	-	-	1
REVI	SNV	2:100040681	G	T	14,40%	Exon 10	Missense	p.Gln537Lys	-	-	1
REVI	SNV	2:100022868	A	T	9,10%	Exon 16	Missense	p.Phe845Ile	-	-	1
REVI	SNV	2:100024509	A	T	8,90%	Exon 15	Missense	p.Asn810Lys	-	-	1
REVI	SNV	2:100022419	C	T	8,50%	Exon 17	Missense	p.Val922Met	-	-	1
REVI	SNP	2:100020945	C	T	50,70%	Exon 18	Missense	p.Ala1003Thr	rs3087401	COSM5114887	1
REVI	SNV	2:100022538	G	A	9,90%	Exon 17	Missense	p.Ser882Leu	-	-	1
REVI	SNV	2:100024550	C	T	7,50%	Exon 15	Missense	p.Gly797Arg	-	-	1
REVI	Indel	2:100024512	TA	T		Exon 15	LoF - (Frameshift)	p.Asn810Ilefs	-	-	1
					5,70%			*5			
REVI	Indel	2:100040638	TG	T		Exon 10	LoF - (Frameshift)	p.Gln551Lysf	-	-	1
					5,50%			s*24			
REVI	SNV	2:100050807	C	T	8,40%	Exon 8	Sinônima	p.Leu475=	-	-	1
REVI	SNP	2:100029250	A	T	12,10%	Exon 13	Sinônima	p.Thr705=	rs756401282	-	1
REVI	SNP	2:100019216	C	T	10,90%	Exon 21	Sinônima	p.Leu1144=	rs1367690549	-	1
REVI	SNV	2:100029196	A	C	11,10%	Exon 13	Sinônima	p.Thr723=	-	-	1
REVI	SNV	2:100050816	A	G	5,40%	Exon 8	Sinônima	p.Asn472=	-	-	1
REVI	SNV	2:100027384	A	T	36%	Intron 13	Variante intron	p.?	-	-	1
REVI	SNV	2:100020310	T	A	5%	Intron 18	Variante intron	p.?	-	-	1

REVI	SNP	2:100027444	C	A	7,80%	Intron 13	Variante intron	p.?	rs1677058100	-	1
REVI	SNV	2:100052295	C	T	5,40%	Intron 7	Variante doador de splice	p.?	-	-	1
REVI	SNV	2:100050729	C	T	5,50%	Intron 8	Variante intron	p.?	-	-	1
REVI	SNP	2:100020292	C	T	47%	Intron 18	Variante intron	p.?	rs138292917	-	1
REVI	SNP	2:100018858	A	G	22,60%	Intron 21	Variante intron	p.?	rs1015664521	-	1
REVI	Indel	2:100020297	TC	T	7,40%	Intron 18	Variante intron	p.?	-	-	1
REVI	Indel	2:100035441	TCA	T	74,90%	Intron 11	Variante intron	p.?	rs201548583	-	1
REVI	Indel	2:100022353	TA	T	16,30%	Intron 17	Variante intron	p.?	-	-	1

Alelo Ref.: alelo referência; Alelo alt.: alelo alterado; VAF: *Variant allele frequency*, frequência do alelo variante; Consequência: consequência da variante para a proteína. Mudança de AA: mudança de aminoácido resultante; dbSNP ID: número de referência da variante no banco de dados dbSNP do NCBI; COSMIC ID: número de referência da variante na plataforma COSMIC. N Amostras detectadas: número de amostras nas quais a variante foi detectada no presente estudo. *LoF: Loss of function*, variantes de perda de função. Células contendo “-“ referem-se a informações que não constam nos bancos de dados em questão.

Fonte: Elaborada pela autora.

Figura 9 - Caracterização das variantes do gene *REVI* detectadas nos pacientes com SMD.



**7A.** Distribuição do número de variantes detectadas para cada éxon referente à região codificadora de *REVI*. **7D.** Classificação das variantes detectadas na região exônica de *REVI*: *missense* (troca de aminoácido); *sinônima* (sem troca de aminoácido); *frameshift* (mudança no quadro de leitura); *stop gained* (códon de parada prematura). **7C.** Distribuição do número de variantes detectadas em cada íntron referente à região não codificadora de *REVI*. **7B.** Classificação das variantes detectadas na região intrônica de *REVI*: variante íntron simples (sem consequência aparente); variante *splice donor* (situada em sítio doador de splice); deleção (remoção de uma ou mais bases).

#### ***4.1.2 Análise in silico da patogenicidade das variantes identificadas na região exônica do gene REVI***

Em relação à análise *in silico* das variantes de nucleotídeo único do tipo *missense* identificadas em *REVI*, 10 variantes foram consideradas como patogênicas, adotando o critério de score mínimo 2, ou seja, no mínimo dois dos quatro algoritmos utilizados (PROVEAN, SIFT, MutationTaster e MutPred) concordaram entre si acerca do potencial patogênico da mutação. Obteve-se 5 variantes de escore máximo (4) e dentre as variantes patogênicas restantes, 4 obtiveram escore 3 e 1 obteve escore 2 (Tabela 7). Além disso, nenhuma dessas variantes classificadas como patogênicas havia sido descrita anteriormente na literatura, tampouco possuem registro vinculado a indivíduos com SMD, portanto, estão relatadas a seguir em mais detalhes acerca das variáveis clínicas dos pacientes em que foram detectadas.

A variante de consequência *missense* p.Glu571Asp, localizada no éxon 11, foi detectada em todos os 14 pacientes incluídos em nosso estudo e teve o seu efeito prejudicial previsto por todos os algoritmos incluídos na análise *in silico*, obtendo escore máximo (4), representando assim um potencial marcador para SMD em ambos os estágios (inicial e avançado).

A variante *missense* p.Phe1229Leu, localizada próximo à região final do gene *REVI*, no éxon 23, obteve escore preditivo de patogenicidade 4, e foi detectada em 1 paciente de baixo risco, diagnosticado com SMD-SA-DM, cujo cariótipo apresentou deleção do cromossomo Y, além de displasia em 2 linhagens medulares e 1 citopenia.

Já a variante de consequência *missense* p.Ile610Asn, também localizada no éxon 11, a qual obteve escore 4, foi detectada em 1 paciente de alto risco que veio a óbito, diagnosticado com o subtipo SMD EB 2 (estágio avançado), cuja análise de cariótipo permitiu identificar dupla alteração incluindo a deleção 7q, fator determinante de um quadro de prognóstico desfavorável. O paciente apresentou ainda 2 displasias e 3 citopenias.

A variante *missense* p.Ile716Asn, localizada no éxon 13, atingiu escore preditivo de patogenicidade 4, tendo sido detectada em 1 paciente com estágio inicial da doença (SMD-SA-DU), o qual apresentou risco muito baixo e cariótipo com alteração indicativa de prognóstico favorável (deleção do cromossomo Y).

A variante *missense* p.Thr369Pro, localizada no éxon 6, a qual obteve escore preditivo de patogenicidade 3, foi detectada em 1 paciente com estágio avançado da doença

(SMD-EB 2) que veio a óbito e apresentou 2 displasias e 3 citopenias, com cariótipo indeterminado pela ausência de metáfases encontradas e consequentemente com classificação de risco pelo IPSS-R indeterminada.

A variante *missense* p.Lys749Asn, localizada no éxon 14, foi avaliada como patogênica por 3 algoritmos, tendo sido detectada em 1 paciente com histórico de progressão da doença, diagnosticado inicialmente como SMD com excesso de blastos I que evoluiu para SMD - EB 2, atingindo uma classificação de risco muito alto, cariótipo complexo com múltiplas alterações incluindo deleção do cromossomo 5q, além de 3 displasias e 3 citopenias.

A variante *missense* p.Gln537Lys, localizada no éxon 10, atingiu escore preditivo de patogenicidade mínimo para esta análise (2), e foi detectada em 1 paciente de alto risco que veio à óbito, tendo sido diagnosticado com SMD - EB 2 e apresentado cariótipo com ausência de metáfases, além de 3 citopenias e 2 displasias medulares.

A variante *missense* p.Arg92Thr, localizada no éxon 4, obteve escore máximo (4), e foi detectada em 1 paciente com SMD em estágio inicial (SMD-SA-DM), que apresentou 3 displasias e 2 citopenias, e cuja análise de cariótipo não permitiu identificar alterações cromossômicas em detrimento da ausência de metáfases na amostra coletada.

A variante *missense* p.Met72Ile, também localizada no éxon 4, teve potencial patogênico previsto por 3 dos 4 algoritmos empregados, sendo detectada em apenas 1 paciente, o qual foi diagnosticado SMD-SA-DM (estágio inicial) e apresentou 2 displasias medulares e 1 citopenia no sangue periférico, além de cariótipo sem alterações, indicando um quadro de risco muito baixo.

A variante de consequência *missense* p.Asn1235Lys, localizada no éxon 23, obteve escore 3, detectada em 1 paciente em estágio inicial diagnosticada com SMD-SA-DU, a qual apresentou cariótipo com ausência de metáfases, 1 displasia em linhagem celular e 1 citopenia no sangue periférico.

As variantes do tipo indel de consequência *frameshift* p.Asn810Ilefs\*5, detectada em um paciente com SMD em estágio avançado (SMD-EB 2), e p.Gln551Lysfs\*24, detectada em um paciente com SMD em estágio inicial (SMD-SA-DM), assim como a variante de consequência *stop gained*, p.Glu118\*, detectada em todos os 14 pacientes, todas referentes ao gene *REVI* e classificadas como variantes *LoF* (*loss of function*) em potencial, foram consideradas no presente estudo como claramente patogênicas devido à gravidade da

consequência que trazem para a proteína, uma vez que estas implicam em mudanças drásticas no quadro de leitura e parada prematura da tradução, respectivamente.

Tabela 7 – Análise in silico de predição de patogenicidade das variantes do tipo *missense* referentes ao gene *REVI*.

dbSNP ID	N Amostras detectadas	VAF	Éxon	Mudança de AA	PROVEAN	SIFT	MutationTaster	MutPred	Escore total	Publicações
-	14	6,8%	Exon 11	p.Glu571Asp	Deletéria	Prejudicial	Causadora de Doença	Patogênica	4	0
-	1	5,2%	Exon 11	p.Ile610Asn	Deletéria	Prejudicial	Causadora de Doença	Patogênica	4	0
-	1	9,7%	Exon 13	p.Ile716Asn	Deletéria	Prejudicial	Causadora de Doença	Patogênica	4	0
-	1	30,2%	Exon 23	p.Phe1229Leu	Deletéria	Prejudicial	Causadora de Doença	Patogênica	4	0
-	1	5,4%	Exon 4	p.Arg92Thr	Deletéria	Prejudicial	Causadora de Doença	Patogênica	4	0
rs370682495	1	51,4%	Exon 4	p.Met72Ile	Deletéria	Tolerada	Causadora de Doença	Patogênica	3	0
rs1032393262	1	50,9%	Exon 6	p.Thr369Pro	Neutra	Prejudicial	Causadora de Doença	Patogênica	3	0
-	1	9,9%	Exon 14	p.Lys749Asn	Deletéria	Prejudicial	Causadora de Doença	Benigna	3	0
-	1	23,9%	Exon 23	p.Asn1235Lys	Neutra	Prejudicial	Causadora de Doença	Benigna	3	0
-	1	14,4%	Exon 10	p.Gln537Lys	Neutra	Tolerada	Causadora de Doença	Patogênica	2	0
rs3087399	2	16,2%	Exon 6	p.Asn373Ser	Deletéria	Tolerada	Polimorfismo	Benigna	1	9
-	1	7,5%	Exon 15	p.Gly797Arg	Neutra	Prejudicial	Polimorfismo	Benigna	1	0
rs3087386	9	50,3%	Exon 6	p.Phe257Ser	Neutra	Tolerada	Polimorfismo	Benigna	0	6
rs3087403	3	46,2%	Exon 5	p.Val138Met	Neutra	Tolerada	Polimorfismo	Benigna	0	6
-	2	5,1%	Exon 17	p.Ser921Arg	Neutra	Tolerada	Polimorfismo	Benigna	0	0
-	2	25%	Exon 21	p.Ser1143Ile	Neutra	Tolerada	Polimorfismo	Benigna	0	0
-	1	9,1%	Exon 16	p.Phe845Ile	Neutra	Tolerada	Polimorfismo	Benigna	0	0
-	1	8,9%	Exon 15	p.Asn810Lys	Neutra	Tolerada	Polimorfismo	Benigna	0	0
-	1	8,5%	Exon 17	p.Val922Met	Neutra	Tolerada	Polimorfismo	Benigna	0	0
rs3087401	1	50,7%	Exon 18	p.Ala1003Thr	Neutra	Tolerada	Polimorfismo	Benigna	0	0
-	1	9,9%	Exon 17	p.Ser882Leu	Neutra	Tolerada	Polimorfismo	Benigna	0	0

dbSNP ID: número de referência da variante no banco de dados dbSNP do NCBI; VAF: *Variant allele frequency* - frequência do alelo variante; Mudança de AA: mudança de aminoácido resultante; N Amostras detectadas: número de amostras nas quais a variante foi detectada no presente estudo; PROVEAN/SIFT/MutationTaster/MutPred: resultado da análise de patogenicidade *in silico* das variantes de acordo com cada ferramenta. Escore total: número total de ferramentas que concordaram entre si acerca do potencial patogênico da variante. Publicações: número de referências encontradas para determinada variante na literatura. Células contendo “-“ referem-se a informações que não constam nos bancos de dados em questão.

Fonte: Elaborada pela autora.

Determinadas variantes exônicas que já haviam sido depositadas em bancos de dados genômicos e classificadas previamente como SNPs foram detectadas em um número relevante dentre os pacientes incluídos em nosso estudo, contudo, o escore total obtido para esses polimorfismos revelou um baixo potencial deletério, apesar dos relatos existentes na literatura demonstrando suas associações a outros tipos de câncer, que serão discutidos posteriormente. As informações acerca do quadro clínico dos pacientes nos quais esses SNPs foram detectados em nosso estudo estão descritos a seguir.

O SNP rs3087386 (p.Phe257Ser), localizado no éxon 6, o qual obteve escore total 0 para previsão de patogenicidade *in silico*, foi detectado em 9 pacientes, sendo estes majoritariamente (5) indivíduos diagnosticados em estágio inicial da SMD, que apresentaram cariótipos normais, enquanto os demais eram pacientes em estágio avançado, dos quais 3 vieram a óbito, sendo estes portadores de cariótipos caracterizados por múltiplas aneuploidias, além disso, um deles possuía cariótipo contendo del(7q).

O SNP rs3087403 (p.Val138Met), localizado no éxon 5, o qual não demonstrou potencial patogênico de acordo com a análise empregada (escore 0), foi detectado em 3 pacientes em nosso estudo, sendo 2 deles diagnosticados em estágio avançado (SMD EB 1 e SMD EB 2), ambos de alto risco, e 1 paciente com SMD SA-DU com cariótipo alterado com deleção do cromossomo Y isolada.

O SNP rs3087399 (p.Asn373Ser), localizado no éxon 6, no presente estudo não obteve o escore de patogenicidade mínimo e foi detectado em 2 pacientes, sendo um deles diagnosticados com SMD SA-DM de baixo risco. O outro paciente, diagnosticado em estágio avançado (SMD- EB 1), apresentou cariótipo com aneuploidias adicionais a del(7q) e veio a óbito.

#### ***4.1.3 Análise in silico da patogenicidade de variantes identificadas na região intrônica do gene REVI***

Quanto às variantes intrônicas, ao utilizar a ferramenta RegulomeDB para prever os efeitos deletérios das referidas mutações encontradas nas regiões não-codificadoras do gene *REVI* e adotando o escore mínimo previsto a partir de “3a” na ferramenta empregada para essa análise, foi possível identificar 4 variantes patogênicas em potencial, de modo que 3 variantes obtiveram alta probabilidade e 1 variante obteve média probabilidade de afetar a ligação a um alvo genético em detrimento da sua localização específica, foram elas: uma

deleção no íntron 18 (c.3045\_3045delG) com escore 2a e alta probabilidade de efeito deletério, 1 SNP (c.3046-14G>A) e 1 SNV (c.3040A>T), localizados no íntron 18 com escores 2a e 2b de alta probabilidade, respectivamente, e um SNV no íntron 13 (c.2170T>A) com escore 3a e média probabilidade de afetar a ligação, todas sem registros prévios em publicações (Tabela 8).

Tabela 8 – Análise *in silico* das variantes detectadas na região intrônica de *REVI*.

dbSNP ID	N Amostras detectadas	VAF	Íntron	cDNA	Escore Regulo meDB	Prob. efeito deletério	Publi cações
-	1	5%	Intron 18	c.3040A>T	2b	<b>Alta</b>	0
rs138292917	1	47%	Intron 18	c.3046-14G>A	2a	<b>Alta</b>	0
-	1	7,40%	Intron 18	c.3045_3045del G	2a	<b>Alta</b>	0
-	1	36%	Intron 13	c.2170T>A	3a	<b>Média</b>	0
rs1677058100	1	7,80%	Intron 13	c.2173-115G>T	4	Muito baixa	0
-	10	5,80%	Intron 13	c.2171T>G	4	Muito baixa	0
rs13394927	3	20,40%	Intron 16	c.2615-49C>T	5	Muito baixa	0
rs13409359	2	30,10%	Intron 8	c.1438+54T>G	5	Muito baixa	0
rs717454	9	41,90%	Intron 16	c.2614+15A>G	5	Muito baixa	2
rs10175852	2	10,30%	Intron 6	c.1213+30G>T	5	Muito baixa	0
-	1	5,50%	Intron 8	c.1438 G>A	5	Muito baixa	0
rs1015664521	1	22,60%	Intron 21	c.3542-49T>C	5	Muito baixa	0
-	1	16,30%	Intron 17	c.2811_2811del T	5	Muito baixa	0
rs28369944	3	41,90%	Intron 2	c.54+67A>G	6	Muito baixa	0
rs28369943	10	54,50%	Intron 2	c.54+61G>A	6	Muito baixa	0
rs201548583	1	74,90%	Intron 11	c.1832- 34_1832-33del	6	Muito baixa	0
-	1	5,40%	Intron 7	c.1319G>A	7	Muito baixa	0

dbSNP ID: número de referência da variante no banco de dados dbSNP do NCBI; VAF: *Varint allele frequency*, frequência do alelo variante; cDNA: alteração de bases na fita complementar; N Amostras detectadas: número de amostras nas quais a variante foi detectada no presente estudo; Escore RegulomeDB: resultado da análise do potencial de patogenicidade das variantes de acordo com a ferramenta; Prob. efeito deletério: probabilidade de determinada variante afetar funções regulatórias de acordo com a ferramenta. Publicações: número de referências encontradas para determinada variante na literatura. Células contendo “-” referem-se a informações que não constam nos bancos de dados em questão.

Fonte: Elaborada pela autora.

A variante c.2170T>A, localizada no íntron 13, a qual obteve escore 3a (média probabilidade), sendo considerada como patogênica em potencial, foi detectada em 1 paciente com estágio avançado da doença (SMD – EB 1) que veio a óbito, tendo apresentado 2 displasias e 3 citopenias, com análise de cariótipo indeterminado em detrimento da ausência de metáfases na amostra coletada.

A variante c.3040A>T, localizada no íntron 18, a qual obteve escore 2b (alta probabilidade), foi detectada em 1 paciente em estágio inicial da doença (SMD-SA-DU), configurando quadro de risco muito baixo, tendo apresentado diseritropoiese e dismegacariocitopenia, além de 1 citopenia no sangue periférico.

O SNP rs138292917 (c.3046-14G>A), localizado também no íntron 18, o qual obteve escore 2a (alta probabilidade) de acordo com a ferramenta utilizada, foi detectado na amostra de 1 paciente em estágio inicial da doença (SMD-SA-DM), que apresentou 2 displasias em achados de biópsia de medula óssea e 1 citopenia no sangue periférico, além de uma análise de cariótipo que não permitiu inferir classificação de risco do seu quadro em razão da ausência de metáfases.

A deleção c.3045\_3045delG, localizada no íntron 18, cujo escore preditivo de patogenicidade alcançou 2a (alta probabilidade), foi detectada em 1 paciente de alto risco de acordo com o IPSS-R, pertencente ao subtipo SMD-EB 2 (estágio avançado) que veio a óbito e apresentou cariótipo anormal com presença de deleção do cromossomo 7q e trissomia do cromossomo 8, indicativo de prognóstico desfavorável, bem como 2 displasias identificadas em linhagens medulares e 3 citopenias no sangue periférico.

O SNP c.2614+15A>G (rs717454), localizado no íntron 16, foi considerado como não patogênico de acordo com nossa análise e foi detectado em 9 pacientes, dentre eles, observou-se majoritariamente em estágios avançados da doença, com prognósticos desfavoráveis e classificação de alto risco, de modo que 3 vieram a óbito. Tal polimorfismo de nucleotídeo único já havia sido reportado na literatura em dois estudos, ocorrência que será discutida posteriormente.

## 4.2 Variantes obtidas para o gene *POLI*

### 4.2.1 Classificação e caracterização das variantes

Após a conclusão do sequenciamento alvo e das análises de processamento iniciais dos dados genômicos extraídos a partir das amostras dos pacientes no presente estudo, foi possível identificar 20 variantes exônicas e 8 variantes intrônicas para o gene *POLI*, totalizando 28 variantes para esse gene. A distribuição das variantes encontradas entre os 10 éxons presentes na sequência codificadora de *POLI* revelou uma maior concentração de mutações próxima à região *downstream* do gene, sendo o éxon 10 aquele com a maior frequência de mutações. Já para na distribuição das variantes pertencentes à região intrônica foi possível observar uma maior frequência de mutações no íntron 9 (Figura 10).

Dentre as variantes exônicas, identificaram-se 18 SNVs, classificados de acordo com sua consequência para a proteína: 13 variantes pontuais *missense* e 4 variantes pontuais sinônimas, além de 1 variante pontual *stop gained* (*LoF*). Ademais, outras 2 variantes do tipo indel foram identificadas, sendo 1 deleção de consequência *frameshift* (mudança no quadro de leitura) e 1 deleção *in frame*. Oito das variantes exônicas referentes ao gene *POLI* identificadas nos pacientes com SMD no presente estudo já haviam sido registradas como polimorfismos de nucleotídeo único – SNPs, possuindo registros do banco de dados dbSNP: rs770802279, rs8305, rs202045769, rs2088038575, do tipo *missense*; rs1324576271, rs376396991, rs3730669, do tipo sinônimas; e rs10584411, uma deleção *in frame*. Por fim, 1 das variantes também já havia sido reportada na plataforma COSMIC (COSM3692146). As demais variantes exônicas encontradas no gene *POLI* ainda não possuem registros nas plataformas mencionadas (Tabela 9).

As variantes identificadas em regiões não-codificadoras do referido gene, em sua grande maioria (6), foram classificadas como variantes intrônicas pontuais simples, com exceção de 2 delas, as quais configuravam deleção de uma ou mais bases. Além disso, 7 já haviam sido registradas como variantes comuns na população de acordo com o banco de dados dbSNP, constando os números de identificação sucedentes: rs660793, rs610281, rs3730688, rs3730809, rs34362008, rs3730727 e rs3730689. Por fim, 3 das variantes intrônicas já constavam na plataforma COSMIC (COSN15304847, COSN6670344, COSN8867434). As informações acerca da classificação, localização, tipo, consequência resultante para a proteína, registros encontrados nos bancos de dados e quantidade de

amostras nas quais as variantes mencionadas foram detectadas estão dispostas em mais detalhes na Tabela 9.

Tabela 9 – Variantes identificadas no gene *POLI* referentes às amostras dos pacientes diagnosticados com SMD.

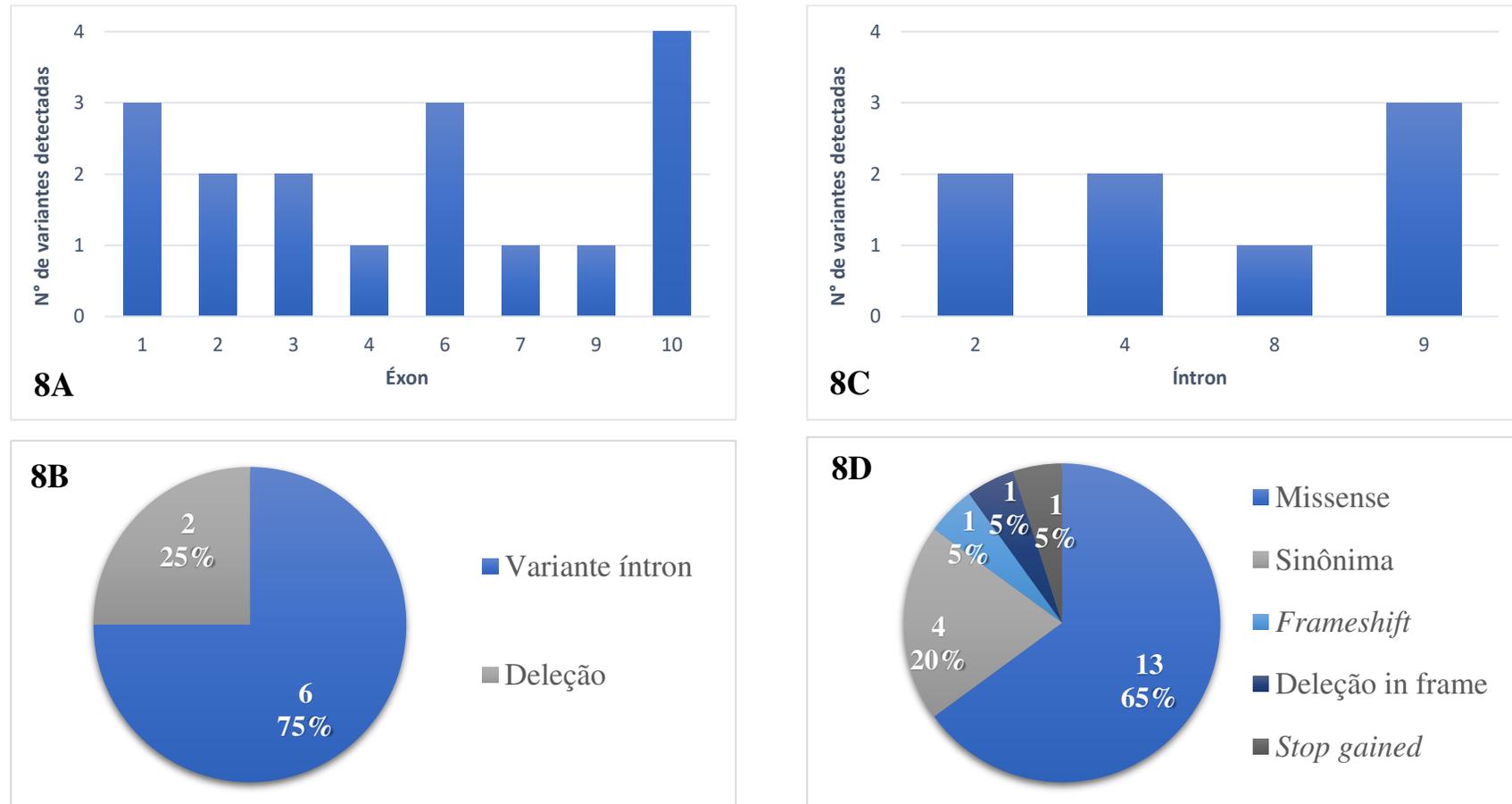
Gene	Tipo de variante	Posição genômica (GRCh37/hg19)	Alelo Ref.	Alelo Alt.	VAF	Localização	Consequência	Mudança de AA	dbSNP ID	COSMIC ID	N Amostras detectadas
<i>POLI</i>	SNV	18:51804193	T	G	11,6%	Éxon 4	<i>Missense</i>	p.Val176Gly	-	-	14
<i>POLI</i>	SNV	18:51820549	C	A	19%	Éxon 10	<i>Missense</i>	p.His645Gln	-	-	12
<i>POLI</i>	Indel	18:51795957	GCG A	G	37,2%	Éxon 1	Deleção <i>in frame</i>	p.Asp17del	rs10584411	COSM288348	10
<i>POLI</i>	SNV	18:51800420	A	C	5,6%	Éxon 3	<i>Missense</i>	p.Glu122Asp	-	-	9
<i>POLI</i>	SNP	18:51819989	G	C	49%	Intron 9	Variante intron	p.?	rs660793	COSM8866668	6
<i>POLI</i>	SNP	18:51804259	C	T	52,9%	Intron 4	Variante intron	p.?	rs610281	COSM8866667	6
<i>POLI</i>	SNV	18:51809327	T	G	5,1%	Éxon 6	<i>Missense</i>	p.Ile306Ser	-	-	6
<i>POLI</i>	SNP	18:51820805	G	A	43%	Éxon 10	<i>Missense</i>	p.Ala731Thr	rs8305	-	5
<i>POLI</i>	SNP	18:51797948	A	G	51,1%	Intron 2	Variante intron	p.?	rs3730688	-	3
<i>POLI</i>	Indel	18:51818417	ACT	A	41,2%	Íntron 9	Variante intron	p.?	rs3730809	-	2
<i>POLI</i>	Indel	18:51818448	CTT	C	43%	Intron 9	Variante intron	p.?	rs34362008	-	2
<i>POLI</i>	SNV	18:51820232	C	T	6,4%	Éxon 10	<i>LoF – (Stop gained)</i>	p.Gln540*	-	COSM4275745	1
<i>POLI</i>	Indel	18:51800401	TG	T	7,3%	Éxon 3	<i>LoF – (Frameshift)</i>	p.Val117Tyrfs *2	-	-	1
<i>POLI</i>	SNV	18:51809321	A	C	5,1%	Éxon 6	<i>Missense</i>	p.Gln304Pro	-	-	1
<i>POLI</i>	SNV	18:51820242	T	C	7,20%	Éxon 10	<i>Missense</i>	p.Ile543Thr	-	-	1
<i>POLI</i>	SNV	18:51820606	T	A	5,1%	Éxon 10	<i>Missense</i>	p.Ser664Arg	-	-	1
<i>POLI</i>	SNV	18:51809295	A	T	5,7%	Éxon 6	<i>Missense</i>	p.Glu295Asp	-	-	1

<i>POLI</i>	SNP	18:51797763	C	T	10,9%	Éxon 2	<i>Missense</i>	p.Ala50Val	rs770802279	-	1
<i>POLI</i>	SNV	18:51820548	A	C	9,2%	Éxon 10	<i>Missense</i>	p.His645Pro	-	-	1
<i>POLI</i>	SNP	18:51795990	C	G	100%	Éxon 1	<i>Missense</i>	p.Ala25Gly	rs202045769	-	1
<i>POLI</i>	SNP	18:51818367	C	T		Éxon 9	<i>Missense</i>	p.Pro455Ser	rs208803857	-	1
					5%				5		
<i>POLI</i>	SNP	18:51810294	C	T	5,2%	Éxon 7	Sinônima		rs132457627	-	1
								p.Ser326=	1		
<i>POLI</i>	SNP	18:51796009	G	C	5,8%	Éxon 1	Sinônima	p.Val31=	rs376396991	-	1
<i>POLI</i>	SNP	18:51796003	G	A	37,2%	Éxon 2	Sinônima	p.Ala29=	rs3730669	COSM3692146	1
<i>POLI</i>	SNV	18:51820207	C	T	11,3%	Éxon 10	Sinônima	p.Val531=	-	-	1
<i>POLI</i>	SNP	18:51804317	A	G	30,8%	Intron 4	Variante intron	p.?	rs3730727	-	1
<i>POLI</i>	SNV	18:51813817	T	C	6%	Intron 8	Variante intron	p.?	-	-	1
<i>POLI</i>	SNP	18:51797951	G	A	59,8%	Intron 2	Variante intron	p.?	rs3730689	-	1

Alelo Ref.: alelo referência; Alelo alt.: alelo alterado; VAF: *Variant allele frequency*, frequência de alelo variante; Mudança de AA: mudança de aminoácido correspondente à variante; dbSNP ID: número de referência da variante no banco de dados dbSNP do NCBI; COSMIC ID: número de referência da variante na plataforma COSMIC. *LoF*: *Loss of function*, variantes de perda de função. Células contendo “-” referem-se a informações que não constam nos bancos de dados em questão.

Fonte: Elaborada pela autora.

Figura 10 – Caracterização das variantes do gene *POLI* detectadas nos pacientes com SMD.



**8A.** Distribuição do número de variantes detectadas para cada éxon referente à região codificadora de *POLI*. **8D.** Classificação das variantes detectadas na região exônica de *POLI*: *missense* (troca de aminoácido); *sinônima* (sem troca de aminoácido); *frameshift* (mudança no quadro de leitura); *deleção in frame* (remoção de uma ou mais bases sem alteração do quadro de leitura); *stop gained* (códon de parada prematura). **8C.** Distribuição do número de variantes detectadas em cada íntron referente à região não codificadora de *POLI*. **8B.** Classificação das variantes detectadas na região intrônica de *POLI*: variante íntron simples (sem consequência aparente); *deleção* (remoção de uma ou mais bases).

#### ***4.2.2 Análise in silico da patogenicidade das variantes identificadas na região exônica de POLI***

Em relação à análise *in silico* das variantes de nucleotídeo único do tipo *missense* referentes ao gene *POLI*, 6 variantes foram consideradas como patogênicas, adotando o critério de score mínimo 2, ou seja, no mínimo dois dos quatro algoritmos utilizados (PROVEAN, SIFT, MutationTaster e MutPred) concordaram entre si acerca do potencial patogênico da mutação. Obteve-se 3 variantes de escore máximo (4), uma delas, p.Glu122Asp localizada no éxon 3, foi detectada em 9 amostras dos pacientes com SMD e teve o seu efeito prejudicial previsto por todos os algoritmos incluídos no estudo. Dentre as variantes patogênicas restantes, 3 obtiveram escore 2, sendo detectadas cada uma em apenas um paciente (Tabela 10). Não foi possível encontrar na literatura estudos destacando tais variantes em associação a quaisquer doenças, sendo esta, portanto, a primeira descrição relatando a sua presença em indivíduos com SMD.

A variante *missense* p.Glu122Asp, localizada no éxon 3, a qual foi considerada como patogênica por obter escore máximo preditivo no presente estudo, foi detectada em 9 pacientes, dentre eles, 5 eram pacientes em estágio avançado (SMD - EB 1 e SMD - EB 2), com quadros clínicos de alto risco de acordo com o IPSS-R, caracterizados por histórico de progressão da doença e eventualmente óbito, cariótipos com alterações cromossômicas de prognóstico desfavorável (duplas aneuploidias, 7q- e trissomia do cromossomo 8), enquanto os demais 4 pacientes estavam em estágio inicial da SMD (SMD-SA-DU e SMD-SA-DM).

Já a variante *missense* p.Ile306Ser, localizada no éxon 3, a qual obteve escore 4 de acordo com os algoritmos utilizados e, por conseguinte, foi considerada como patogênica, foi detectada em 6 indivíduos em nosso estudo, os quais 2 eram pacientes em estágio avançado de SMD, de alto risco e com prognósticos desfavoráveis que vieram à óbito, enquanto os demais estavam em estágio inicial da doença, com prognósticos favoráveis e quadros de baixo risco.

Em relação à variante do tipo *missense* p.Gln304Pro, também localizada no éxon 6, com escore preditivo de patogenicidade 4 obtido a partir das ferramentas de análise *in silico*, foi detectada em 1 paciente com SMD em estágio avançado (SMD - EB 1), apresentando um quadro de alto risco segundo os critérios do IPSS-R e cariótipo sem alterações cromossômicas, além de 3 citopenias no sangue periférico e amostra medular caracterizada por 2 displasias (diseritropoiese e disgranulopoiese).

Tabela 10 - Análise *in silico* das variantes do tipo *missense* detectadas para o gene *POLI*.

dbSNP ID	N Amostras detectadas	VAF	Éxon	Mudança de AA	PROVEAN	SIFT	MutationTaster	MutPred	Escore total	Publicações
-	6	5,1%	Éxon 6	p.Ile306Ser	Deletéria	Prejudicial	Causadora de Doença	Patogênica	4	0
-	9	5,6%	Éxon 3	p.Glu122Asp	Deletéria	Prejudicial	Causadora de Doença	Patogênica	4	0
-	1	5,1%	Éxon 6	p.Gln304Pro	Deletéria	Prejudicial	Causadora de Doença	Patogênica	4	0
-	1	7,20%	Éxon 10	p.Ile543Thr	Neutra	Prejudicial	Causadora de Doença	Benigna	2	0
-	1	5,1%	Éxon 10	p.Ser664Arg	Neutra	Prejudicial	Causadora de Doença	Benigna	2	0
-	1	5,7%	Éxon 6	p.Glu295Asp	Neutra	Prejudicial	Causadora de Doença	Benigna	2	0
rs770802279	1	10,9%	Éxon 2	p.Ala50Val	Neutra	Prejudicial	Polimorfismo	Benigna	1	0
-	1	9,2%	Éxon 10	p.His645Pro	Neutra	Prejudicial	Polimorfismo	Benigna	1	0
-	14	11,6%	Éxon 4	p.Val176Gly	Neutra	Tolerada	Polimorfismo	Benigna	0	0
-	12	19%	Éxon 10	p.His645Gln	Neutra	Tolerada	Polimorfismo	Benigna	0	0
rs8305	5	43%	Éxon 10	p.Ala731Thr	Neutra	Tolerada	Polimorfismo	Benigna	0	10
rs202045769	1	100%	Éxon 1	p.Ala25Gly	Neutra	Tolerada	Polimorfismo	Benigna	0	0
rs2088038575	1	5%	Éxon 9	p.Pro455Ser	Neutra	Tolerada	Polimorfismo	Benigna	0	0

dbSNP ID: número de referência da variante no banco de dados dbSNP do NCBI; VAF: *Variant allele frequency* - frequência do alelo variante; Mudança de AA: mudança de aminoácido resultante; N Amostras detectadas: número de amostras nas quais a variante foi detectada no presente estudo; PROVEAN/SIFT/MutationTaster/MutPred: resultado da análise de patogenicidade *in silico* das variantes de acordo com cada ferramenta. Escore total: número total de ferramentas que concordaram entre si acerca do potencial patogênico da variante. Publicações: número de referências encontradas para determinada variante na literatura. Células contendo “-“ referem-se a informações que não constam nos bancos de dados em questão.

Fonte: Elaborada pela autora.

A variante p.Ile543Thr de consequência *missense*, localizada no éxon 10, cujo escore preditivo de patogenicidade alcançou a pontuação mínima (2) para ser considerada patogênica no presente estudo, foi detectada em 1 paciente em estágio avançado da doença (SMD EB - 1), com um quadro de alto risco, caracterizado pela progressão ao óbito e cariótipo complexo de prognóstico desfavorável, 3 displasias medulares e 2 citopenias.

As variantes de consequência *missense* p.Ser664Arg e p.Glu295Asp, localizadas nos éxons 10 e 6, ambas consideradas patogênicas de acordo com 2 algoritmos utilizados em nossa análise, foram detectadas no mesmo paciente, o qual foi classificado clinicamente como subtipo SMD EB 2 (avançado) e veio a óbito, cuja análise de cariótipo permitiu identificar alterações comumente associadas a prognósticos desfavoráveis (duplas alterações incluindo trissomia do cromossomo 8 e 7q-), medula óssea acometida por 2 anormalidades displásicas (diseritropoiese e dismegacariocitopoiese), além de 3 citopenias no sangue periférico, apresentando um quadro de alto risco segundo o IPSS-R.

No que concerne à avaliação da patogenicidade das variantes cujas consequências são seriamente prejudiciais para a proteína resultante (variantes potencialmente *loss of function*, perda de função) obtidas para o gene *POLI* a partir das amostras dos pacientes diagnosticados em nosso estudo, a variante do tipo indel de consequência *frameshift* (p.Val117Tyrfs\*2), assim como o SNV de consequência *stop gained* (p.Gln540\*), ambos detectados cada um em um paciente com SMD em estágio inicial (SMD-SA-DM e SMD-SA-DU, respectivamente), foram consideradas como claramente patogênicas devido à gravidade da consequência que trazem, pois acarretam na perda de função proteica às custas de mudanças drásticas no quadro de leitura e parada prematura da tradução.

Já a deleção *in frame* p.Asp17delCGA (rs10584411; COSM288348), localização no éxon 1, foi detectada em 10 pacientes no presente estudo, sendo estes predominantemente (6) de baixo risco, com prognósticos favoráveis, referentes aos subtipos SMD-SA-DM (3) e SMD-SA-DU (2). Ademais, foi possível encontrar relatos prévios na literatura sobre esse polimorfismo, que serão discutidos posteriormente.

Um polimorfismo detectado em 7 pacientes incluídos em nossa análise, (rs8305, p.Ala731Thr), localizado no éxon 10, não obteve previsão de patogenicidade de acordo com nenhum dos quatro algoritmos empregados, mas já possuía descrições em publicações anteriores, as quais serão abordadas mais adiante. Em nossos achados, o referido SNP foi identificado majoritariamente em pacientes com SMD em estágio inicial e quadros clínicos de baixo risco.

### ***4.2.3 Análise in silico da patogenicidade das variantes identificadas na região intrônica do gene POLI***

Em relação às variantes intrônicas encontradas para o gene *POLI*, de acordo com os resultados obtidos na ferramenta RegulomeDB, que leva em consideração informações de bancos de dados públicos acerca do potencial regulatório das variantes não codificadoras na expressão de alvos genéticos, a grande parte dos SNPs detectados em nossos pacientes obteve um escore resultante de efeito deletério quase nulo, isto é, a probabilidade dessas mutações afetarem a ligação a um alvo genético específico foi avaliada como muito baixa pelo algoritmo utilizado (Tabela 11).

Entretanto, foi possível identificar uma variante intrônica patogênica em potencial (c.484+92A>G - rs3730727), localizada no íntron 4, a qual obteve um escore 3a (média probabilidade de efeito deletério). Tal SNP não possui relato na literatura, e em nosso estudo foi detectado em apenas 1 paciente em estágio avançado da doença (SMD-EB 2), que manifestou histórico de progressão de SMD com excesso de blastos 1 para SMD com excesso de blastos 2, apresentando cariótipo com múltiplas aneuploidias (>3) incluindo del (5q), alterações cromossômicas associadas a prognóstico desfavorável, além de 3 anormalidades displásicas medulares e 3 citopenias no sangue periférico, variáveis clínicas representativas de um fenótipo de alto risco.

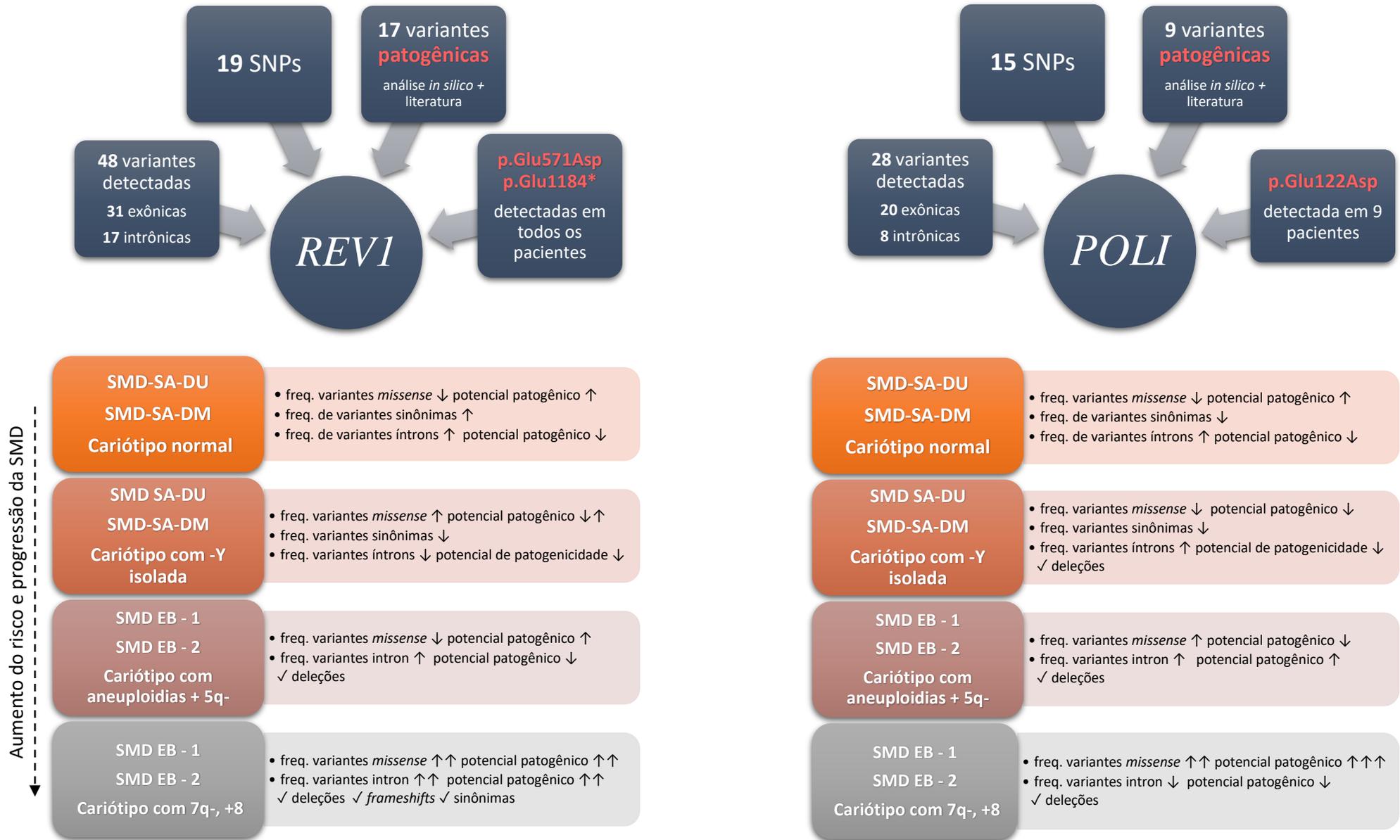
Tabela 11 – Análise *in silico* das variantes intrônicas detectadas no gene *POLI*

dbSNP ID	N Amostras detectadas	VAF	Íntron	cDNA	Escore Regulo meDB	Prob. efeito deletério	Publi cações
rs3730727	1	30,8%	Intron 4	c.484+92A>G	3a	<b>Média</b>	0
rs660793	6	49%	Intron 9	c.140530G>C	6	Muito baixa	0
rs610281	6	52,9%	Intron 4	c.559+34C>T	4	Muito baixa	0
rs3730688	3	51,1%	Intron 2	c.166+93A>G	5	Muito baixa	0
rs3730809	2	41,2%	Íntron 9	c.1329+12_13 29+13delCT	5	Muito baixa	0
rs34362008	2	43%	Intron 9	c.1329+41_13 29+42delTT	6	Muito baixa	0
-	1	6%	Intron 8	c.1127A>G	5	Muito baixa	0
rs3730689	1	59,8%	Intron 2	c.166+96G>A	5	Muito baixa	0

dbSNP ID: número de referência da variante no banco de dados dbSNP do NCBI; VAF: *Variant allele frequency* – frequência do alelo variante; cDNA: alteração de bases na fita complementar; N Amostras detectadas: número de amostras nas quais a variante foi detectada no presente estudo; Escore RegulomeDB: resultado da análise do potencial de patogenicidade das variantes de acordo com a ferramenta. Prob. efeito deletério: probabilidade de determinada variante afetar funções regulatórias de acordo com a ferramenta. Publicações: número de referências encontradas para determinada variante na literatura. Células contendo “-“ referem-se a informações que não constam nos bancos de dados em questão.

Fonte: Elaborada pela autora

Figura 11 – Esquematisação dos principais resultados obtidos em nosso estudo.



## 5 DISCUSSÃO

As células eucarióticas estão continuamente submetidas a danos causados ao DNA, sejam estes causados por meio de exposição a fontes endógenas ou exógenas. A preservação da integridade genômica é um fator essencial manutenção da homeostase celular, de modo que os organismos dispõem de diversas estratégias para suportar estresses genotóxicos, capacitando-os para reparar esses erros de forma eficiente, garantindo a fidelidade do processo replicativo necessária para a sobrevivência celular e evolução dos organismos multicelulares (MAIORANO et al., 2021).

Os mecanismos de resposta ao dano do DNA são responsáveis por reparar os erros identificados, e operam tanto em lesões de fita dupla quanto de fita simples. Entretanto, ao encontrar lesões no DNA específicas mais robustas, as DNA polimerases clássicas paralisam sua atividade por serem incapazes de adicionar nucleotídeos em oposição a tais falhas, comprometendo assim o prosseguimento do processo replicativo e bloqueando a forquilha de replicação, o que pode causar instabilidade genômica, um fenômeno notavelmente associado ao surgimento de mutações aleatórias e alterações cromossômicas que configura uma das características facilitadoras para a aquisição dos dez principais marcadores do câncer, conforme definido e recentemente revisado por Hanahan em seu artigo “*Hallmarks of Cancer: New dimensions*” (HANAHAN, 2022).

Para evitar esses eventos desvantajosos, as células evoluíram para desenvolver mecanismos como a síntese translesão, a qual permite a continuidade do processo replicativo pela ação transiente de DNA polimerases especializadas, capazes de perpassar através e além de lesões volumosas, evitando assim o bloqueio da forquilha de replicação e, por conseguinte, a ocorrência de quebras de dupla fita que eventualmente suscitariam alterações a nível cromossômico, como rearranjos, translocações e deleções (LANGE et al., 2013). Duas das enzimas envolvidas na síntese translesão, bem como seus respectivos genes codificadores, *REVI* e *POLI*, constituem o tema central do presente estudo.

### **5.1 Variantes patogênicas de *REVI* e *POLI* detectadas em um alto número de pacientes representam biomarcadores em potencial para SMD**

Tendo em vista a alta incidência de mutações e alterações citogenéticas encontradas em pacientes com SMD, há evidências crescentes que a maquinaria de reparo do DNA é deficiente na Síndrome Mielodisplásica, denotando a incapacidade das células de lidar com danos no DNA, em um estudo realizado por Ribeiro e colaboradores, por exemplo, polimorfismos de nucleotídeo único referentes ao gene *ATM*, constituinte da via de reparo do DNA responsável pela sinalização e detecção de quebras de dupla fita (DSBs), foram associados significativamente a casos de SMD de baixo risco (RIBEIRO et al., 2013). Destaca-se ainda nesse contexto a relevância da instabilidade genômica no surgimento de múltiplas mutações para a desdobramento do processo patológico em SMD (YU et al., 2020).

Em decorrência dessa desregulação no sistema de reparo, a ineficiência em identificar, sinalizar e reparar adequadamente determinadas lesões do DNA leva células danificadas a recorrerem a outras estratégias, como a via de tolerância ao dano, para lidar com as lesões existentes sem comprometer o processo replicativo, numa espécie de mecanismo compensatório desenvolvido no intuito de evitar a morte celular, mesmo às custas de uma redução na fidelidade replicativa (CECCALDI et al., 2015). Ressalta-se que essa estratégia foi ratificada em ensaios celulares com linhagens *knockout* para genes da via Anemia de Fanconi (FA) – um distúrbio genético associado à insuficiência progressiva da medula óssea e profunda suscetibilidade ao câncer – onde se observou que células deficientes nos genes da via de reparo exibiram acentuados níveis de Pol  $\iota$  (enzima referente ao gene *POLI*) justamente como uma tática de escape contra desbalanços genéticos mais graves que acarretariam a apoptose, demonstrando assim uma dependência do mecanismo de síntese translesão para garantir a sobrevivência celular (WANG et al., 2020).

Atentando-se ao número considerável (26, no total) de variantes patogênicas identificadas nos genes *REVI* e *POLI*, diretamente implicados no mecanismo TLS, para os 14 pacientes incluídos em nosso estudo, que haviam sido previamente diagnosticados com SMD e apresentaram quadros clínicos diversos, evidencia-se o papel da perda da integridade genômica no surgimento dessa malignidade hematológica, porém, aponta-se também a possibilidade dos mecanismos de tolerância ao dano estarem sendo afetados pela instabilidade genômica intrínseca à essa doença, o que representaria um grave problema, pois nessa conjuntura ambas as vias, reparo e tolerância ao dano, que protegem as células de

lesões no DNA, se encontrariam insuficientes em pacientes com SMD, desamparando assim o metabolismo celular das estratégias existentes contra o surgimento de anomalias cromossômicas prejudiciais.

Além disso, dentre os resultados obtidos, destaca-se a presença da variante de consequência *missense* p.Glu571Asp, a qual obteve potencial patogênico máximo (escore 4) previsto pelas ferramentas utilizadas, e da variante de consequência *stop gained* p.Glu1184\* ambas referentes ao gene *REVI* e detectadas em todos os 14 pacientes incluídos em nosso estudo, ressalta-se ainda a variante de consequência *missense* p.Glu122Asp, referente ao gene *POLI*, que também obteve máximo escore preditivo de patogenicidade de acordo com a análise empregada e foi detectada em 9 pacientes, sendo os indivíduos deste grupo diagnosticados com SMD tanto em estágio inicial quanto avançado.

Considerando a ausência de referências na literatura a respeito das variantes encontradas e o seu elevado potencial patogênico previsto *in silico*, aliado ao alto número de pacientes nos quais essas variantes foram detectadas, p.Glu571Asp, p.Glu1184\* correspondentes ao gene *REVI* e p.Glu122Asp, identificada em *POLI*, representam candidatas com grande potencial como novos biomarcadores para Síndrome Mielodisplásica, sendo necessário realizar mais estudos acerca da sua influência na progressão e patogênese da doença.

## **5.2 Acúmulo de mutações patogênicas nos genes *REVI* e *POLI* em pacientes em estágio avançado da SMD com cariótipo del (7q)**

Em nosso estudo foi possível identificar que o paciente diagnosticado com forma avançada da SMD (SMD - EB 2), portador de cariótipo com múltiplas alterações cromossômicas incluindo deleção 7q, apresentou acúmulo de mutações do tipo *missense*, sendo estas caracterizadas por potencial patogênico diverso (alto e baixo), além da presença de mutação do tipo perda de função (*frameshift*), mutações sinônimas e um notável acúmulo de alterações intrônicas, incluindo deleções e variantes com alto potencial de afetar funções regulatórias, conforme o estudo realizado *in silico* para o gene *REVI*. Ademais, o paciente também apresentou alta frequência de variantes *missense* com elevados potenciais patogênicos para o gene *POLI*, em concomitância à uma baixa frequência de variantes na região não codificadora desse gene, porém dotadas de escores preditivos de patogenicidade elevados e presença de deleções em ambas regiões, intrônica e exônica.

Atualmente, a polimerase TLS REV1 é mais notavelmente reconhecida pela sua atividade regulatória. Mais especificamente, essa polimerase atua como uma proteína de *scaffold* para as demais enzimas envolvidas na síntese translesão e possui capacidade de interagir simultaneamente com outros componentes do complexo, diretamente envolvidos no processo de transpassagem ao dano, como PCNA, Rev7 (constituente da POLZ), Pol  $\eta$  e Pol  $\kappa$ , tendo assim um papel indispensável no agrupamento do complexo multi-proteico constituinte do mecanismo TLS, denominado de mutassoma. Tais evidências sugerem que ela facilita não só o recrutamento, como também a regulação das polimerases TLS, de forma similar à PCNA (TISSIER et al., 2004; JAIN et al., 2018; GUO, 2003).

Ressalta-se a função protetora da via de tolerância ao dano na prevenção de eventos mutacionais mais graves, como aberrações cromossômicas desbalanceadas, deleções e rearranjos. Nesse sentido, é possível inferir que a alta frequência de variantes com elevado potencial patogênico encontradas para o paciente de alto risco e prognóstico desfavorável, localizadas tanto na região exônica quanto intrônica de *REV1*, além da presença de mutações mais graves de consequência *frameshift*, somado à alta frequência de variantes com elevado potencial patogênico identificadas em *POLI* para o mesmo paciente, representam um compilado de fatores que possivelmente estariam comprometendo a função regulatória das enzimas resultantes, prejudicando assim o mecanismo TLS como um todo e induzindo eventos como o colapso da forquilha de replicação, que acentuaria a instabilidade genômica nas células e permitiria a geração das alterações cromossômicas observadas no cariótipo em estudo, como a deleção do 7q e trissomia do cromossomo 8.

Além disso, polimorfismos de nucleotídeo único não-sinônimos referentes aos genes *REV1* e *POLI* em estudos anteriores já foram associados significativamente ao aumento do risco de adenocarcinoma e carcinoma de células escamosas em pacientes com câncer de pulmão, denotando assim um histórico prévio de variantes identificadas nesse gene na contribuição para o processo de tumorigênese (SAKIYAMA et al., 2005).

### **5.3 Nulissomia do cromossomo Y e baixa frequência de variantes patogênicas nos genes *REV1* e *POLI***

A perda do cromossomo Y é uma das alterações genômicas somáticas adquiridas mais comuns em homens, estando presente em aproximadamente 6% dos cariótipos de medula óssea masculinos no total e em 16% dos cariótipos medulares alterados. De modo

geral, a nulissomia do cromossomo Y constitui um fator associado à senescência e sua incidência em células da medula aumenta proporcionalmente com a idade, de modo que 20% dos homens saudáveis acima de 80 anos apresentam tal alteração citogenética (OUSEPH; HASSERJIAN, 2021).

Em distúrbios clonais hematopoéticos, a nulissomia do Y é observada de forma mais frequente em contextos onde há presença de outras anomalias cromossômicas, e em SMD configura-se como uma das alterações citogenéticas mais recorrentes, alcançando incidência de até 30% em homens. Em contrapartida, a ocorrência de nulissomia do cromossomo Y isolada em SMD é possui frequência reduzida, com taxas variando de 4% a 10% dentre os cariótipos de pacientes diagnosticados (ZHANG et al., 2007; SHAHRABI et al., 2018). Os indivíduos com SMD portadores dessa alteração citogenética na sua forma isolada costumam alcançar maiores índices de sobrevida global em comparação aos pacientes diagnosticados sem anormalidades cromossômicas, de tal maneira que os cariótipos caracterizados por -Y isolada são classificados de acordo com o IPSS-R como quadros de prognóstico muito favorável (GARCÍA-ISIDORO et al., 1997; GREENBERG et al., 2012).

Os pacientes diagnosticados em estágio inicial da SMD em nosso estudo (subtipos SMD-SA-DU e SMD-SA-DM), cujos cariótipos apresentavam nulissomia isolada do cromossomo Y, manifestaram uma alta frequência de mutações do tipo *missense* para o gene *REVI*, com escores de caráter altos e baixos em relação à previsão de patogenicidade *in silico*, além de uma baixa frequência de mutações sinônimas e baixa frequência de alterações nos íntrons, sendo estas com baixo potencial de patogenicidade. Já para o gene *POLI*, os pacientes pertencentes ao referido grupo apresentaram baixa frequência de mutações do tipo *missense*, com baixo potencial patogênico, baixa frequência de variantes sinônimas e alta frequência de variantes na região intrônica, as quais atingiram baixa probabilidade de afetar funções regulatórias, além da presença de deleções.

A partir desses resultados, tendo em vista que a perda do cromossomo Y constitui um fator associado à senescência que está atrelado a prognósticos favoráveis em SMD, e observando também que os quadros clínicos dos pacientes idosos portadores dessa alteração citogenética eram de baixo risco, conforme categorizado pelo IPSS-R, pode-se inferir que os mecanismos de tolerância ao dano não estão sofrendo maiores prejuízos nesses casos e por isso exercem funcionalmente seu papel protetor contra o surgimento de anomalias cromossômicas, o que é evidenciado pela ausência de aneuploidias ou desarranjos genômicos de maior porte adicionais nos cariótipos obtidos para tais indivíduos. Corroborando com essa

hipótese, ressalta-se a baixa frequência de variantes patogênicas identificadas nos genes constituintes da via de tolerância ao dano dentro desse grupo de pacientes, um indicativo de que a instabilidade genômica associada à SMD não atua como fator disruptivo para a maquinaria de síntese translesão em casos de baixo risco e perda isolada do cromossomo Y, estando esta via ainda ativada e funcional nesses pacientes.

#### **5.4 Pacientes em estágio avançado da SMD e cariótipo alterado com del(5q) apresentam perfil genômico misto para os genes *REVI* e *POLI***

De acordo com o sistema utilizado de forma mais abrangente para a estratificação prognóstica em SMD, o IPSS-R, pacientes portadores de cariótipos com deleção do 5q (ocorrência também chamada de síndrome 5q-), possuem um prognóstico relativamente favorável, com baixo risco de progressão para LMA, estando, portanto situados na categoria de baixo risco para esta doença (GIAGOUNIDIS et al., 2006). Tais pacientes apresentam uma deleção no cromossomo 5, localizada entre as bandas 5q31 e 5q33, com porcentagem de blastos na medula óssea abaixo de 5%, e cerca de 50% dos indivíduos responde bem ao tratamento com lenalidomida, um fármaco imunomodulador recomendado nestes casos. Apesar do baixo risco de transformação para um quadro leucêmico, a dependência transfusional característica da síndrome 5q- denota um efeito prejudicial nesses casos, trazendo complicações. Quando observadas, alterações citogenéticas adicionais à deleção 5q também podem ter efeito negativo sobre a progressão da doença e sobrevida dos pacientes (SCHARENBERG et al., 2017; LODÉ et al., 2018).

Além disso, alterações específicas em genes comumente associados ao câncer podem afetar o prognóstico de pacientes portadores de cariótipos com del(5q). Mutações identificadas em *TP53*, por exemplo, um gene um supressor tumoral de destaque em malignidades hematológicas e no câncer em geral, estão associadas ao perfil citogenético del(5q) isolada, porém mais são mais frequentes em cariótipos complexos com com -5/5q-, resultando em prognósticos desfavoráveis, resistência ao tratamento e menor sobrevida (KULASEKARARAJ et al., 2013; KANEKO et al., 1995).

Em relação aos pacientes presentes em nosso estudo que foram diagnosticados com subtipos SMD - EB 1 e SMD - EB 2, configurando um estágio avançado da doença, dois deles vieram a óbito e eram portadores de cariótipos com aneuploidias, dentre eles, um

paciente possuía múltiplas alterações citogenéticas incluindo deleção (5q). Observou-se nesse grupo uma baixa frequência de mutações do tipo *missense* em relação ao gene *REVI*, porém com alto potencial patogênico, alta frequência de mutações intrônicas com baixo potencial patogênico, além de deleções. Quanto ao gene *POLI*, identificou-se uma alta frequência de variantes do tipo *missense*, com potencial patogênico predominantemente baixo, alta frequência de variantes intrônicas, incluindo uma alteração de caráter patogênico elevado, além da presença de deleções, evidenciando assim um perfil genômico heterogêneo para ambos genes, com frequências mutacionais e patogenicidade variáveis.

Tendo em vista a heterogeneidade dos resultados obtidos e a classificação de risco dos pacientes portadores das variantes detectadas, os quais se alternaram entre risco alto a risco muito alto de acordo com o IPSS-R, a presença de variantes patogênicas e deleções nos genes estudados demonstra, embora com frequências variáveis, que a via de síntese translesão pode estar comprometida ou sobrecarregada nesses casos, em decorrência da instabilidade genômica provocada pela inaptação das células em escapar dos danos provocados ao DNA. Seguindo essa hipótese, a falha do mecanismo de tolerância ao dano estaria permitindo o surgimento das anomalias cromossômicas, tais como aneuploidias, observadas nos cariótipos dos referidos pacientes. Nesse contexto, recomenda-se, portanto, que os indivíduos pertencentes a esse grupo que ainda se encontram em vida sejam acompanhados de perto, em detrimento do risco iminente de progressão da doença para um quadro de leucemia mieloide aguda.

Para além desses achados, é possível ainda conjecturar que o paciente portador da deleção 5q que veio a óbito carregava, além das variantes nas sequências genômicas de *REVI* e *POLI* detectadas em nosso estudo, variantes somáticas condutoras (*driver mutations*) em outros genes implicados na consolidação da oncogênese, como mutações no gene supressor tumoral *TP53*, as quais estão intimamente associadas à instabilidade genômica e a cariótipos complexos contendo deleção 5q (REILLY et al., 2019), impactando diretamente no agravamento da SMD em quadros caracterizados por prognóstico desfavorável, classificação de alto risco para progressão da doença à LMA e menores índices de sobrevida.

### **5.5 Pacientes em estágio inicial da SMD e cariótipo normal apresentam menor frequência de variantes para os genes *REVI* e *POLI***

Foi possível identificar que os pacientes em estágio inicial da doença (SMD-SA-DU e SMD-SA-DM), cujos cariótipos não apresentavam alterações cromossômicas, em relação ao gene *REVI* possuíram menor frequência de mutações do tipo *missense*, porém de escores preditivos de patogenicidade altos, além de um acúmulo de mutações sinônimas e intrônicas, com baixo potencial patogênico. Já para o gene *POLI*, o mesmo grupo de pacientes apresentou baixa frequência de variantes do tipo *missense*, porém com elevado potencial patogênico, baixa frequência de variantes sinônimas e alta frequência de variantes intrônicas, com baixo potencial patogênico.

Esses resultados evidenciam que o perfil molecular obtido está em concordância com o prognóstico favorável dos pacientes, uma vez que seus quadros clínicos se configuram como quadros de baixo risco, apresentando número reduzido de citopenias, poucas displasias na medula óssea e baixa porcentagem de blastos. Observa-se que eles não possuem excesso de mutações patogênicas em genes constituintes da via de tolerância ao dano, e que manifestam uma maior quantidade de mutações somente em regiões não codificadoras dos genes *REVI* e *POLI*, que alcançaram um baixo potencial patogênico de acordo com a análise *in silico*, ou carregam ainda variantes que não implicam em mudança de aminoácido (sinônimas), portanto, infere-se que a maquinaria responsável pela via de tolerância ao dano no DNA não está comprometida nesses casos, de modo que as forquilhas de replicação não sofrem colapso, permitindo que o processo replicativo seja retomado mesmo diante de lesões volumosas.

Tal hipótese é corroborada pela ausência de alterações mais graves, como aneuploidias e anomalias citogenéticas, nos cariótipos analisados para esses pacientes. Nesse sentido, compreende-se que a instabilidade genômica intrínseca à SMD não está suficientemente acentuada em casos com cariótipo normal cujo prognóstico é favorável, sem gerar assim maiores prejuízos aos mecanismos de síntese translesão.

## **5.6 Variantes polimórficas detectadas nos genes *REVI* e *POLI* presentes na literatura**

Polimorfismos são definidos como variações nas sequências de DNA que ocorrem em uma determinada população numa frequência  $\geq 1\%$ , envolvendo alterações em um ou mais nucleotídeos. A forma mais comum de variações no genoma humano é representada por polimorfismos de nucleotídeo único (SNPs), os quais, dependendo da sua

localização específica (em genes associados ao reparo do DNA ou regulação do ciclo celular, por exemplo), podem estar relacionados à predisposição ao câncer.

No que concerne o gene *REVI*, que está diretamente envolvido na via de tolerância ao dano do DNA, já constam mencionados na literatura alguns polimorfismos de nucleotídeo único em estudos relacionando-os tanto ao aumento do risco, quanto à redução da mortalidade em diferentes tipos de câncer. Um dos mais reportados nesse contexto é o SNP rs3087386 (p.Phe257Ser), localizado no éxon 6, ele se encontra descrito em 6 estudos distintos, nos quais se identificou sua associação significativa ao aumento da sobrevivência de pacientes com câncer de pulmão (XU et al., 2013), além de uma menor propensão à neutropenia em casos de mesotelioma maligno (GORICAR et al., 2014). Em contrapartida, Sangrajrang e colaboradores também reportaram sua presença ligada a um aumento no risco de câncer de mama (SANGRAJRANG et al., 2011), e seu impacto negativo foi observado ainda na redução das taxas de sobrevivência em pacientes com câncer de cólon expostos a fatores de risco (JIRASKOVA et al., 2018).

Não há, no entanto, relatos em publicações demonstrando sua presença em indivíduos com Síndrome Mielodisplásica, e em nosso estudo foi possível identificar essa referida variante, do tipo *missense*, em 9 pacientes, sendo estes majoritariamente (5) indivíduos diagnosticados em estágio inicial da SMD, que apresentaram cariótipos normais, enquanto os demais eram pacientes em estágio avançado com cariótipos caracterizados por aneuploidias, dos quais 3 vieram a óbito, um deles apresentando del(7q). Apesar do SNP não ter alcançado escore mínimo para ser considerado como patogênico somente com base no efeito resultante à proteína, de acordo com as ferramentas *in silico* utilizadas (escore total 0), infere-se que ele pode ter algum papel no processo de tumorigênese quando aliado a fatores de risco, tal como Jiraskova e colaboradores verificaram em casos de neoplasias malignas colorretais, sendo necessário traçar mais ensaios experimentais para de fato compreender sua atuação nesse contexto.

De forma similar, o SNP rs3087403 (p.Val138Met), detectado na nossa pesquisa em 3 pacientes que foram diagnosticados com SMD EB-1, SMD EB-2, ambos de alto risco, e SMD-SA-DU com baixo risco, respectivamente, também não obteve previsão de patogenicidade a partir de nenhum dos algoritmos empregados, alcançando escore total 0 para esta análise. Localizado no éxon 5, esse polimorfismo já possui registros prévios onde observou-se que sua presença estava associada a uma menor sobrevivência em pacientes com osteosarcoma, além de afetar a eficácia do tratamento com cisplatina nesses casos, com um

aumento da leucopenia e neutropenia nos quadros clínicos investigados (GORICAR et al., 2015; 2016).

O mesmo grupo de pesquisadores investigou ainda a influência do SNP rs3087399 (p.Asn373Ser), um outro polimorfismo localizado no éxon 6, na progressão dos quadros de pacientes com osteosarcoma e mesotelioma maligno, porém não detectaram associação significativa entre a sua presença e redução ou aumento da sobrevivência dos indivíduos (GORICAR et al., 2015; 2014). Nós detectamos essa variante, que foi classificada como patogênica somente por um dos quatro algoritmos incluídos em nosso estudo, em 2 pacientes com quadros clínicos distintos, sendo um deles pertencente ao subtipo SMD-SA-DM, com variáveis de baixo risco, e outro diagnosticado com SMD em estágio avançado (SMD-EB-2) que veio a óbito, portador de cariótipo com múltiplas aneuploidias e del(7q). Todavia, esse SNP foi descrito em uma publicação de cunho experimental como sendo responsável pela relativa facilitação da atividade catalítica de *REVI* (YEOM et al., 2016). Diante desses dados, inferimos que a atuação da enzima REVI em nossos pacientes estava sob influência de várias mutações, e que, em algum dado momento, uma delas (p.Asn373Ser) pode ter causado efeito sobre o seu desempenho catalítico, alterando o funcionamento normal do mecanismo TLS e induzindo eventos mutagênicos.

Sabe-se que os polimorfismos de nucleotídeo único ocorrem com maior frequência em regiões não codificantes, localizadas em íntrons ou regiões intergênicas, porém, tudo depende da localização em que tais variantes se encontram, uma vez que alterações em elementos intimamente relacionados à regulação da expressão gênica, como acentuadores ou silenciadores, podem comprometer mecanismos funcionais regulatórios. Um dos polimorfismos que detectamos em um sítio não codificante relativo ao íntron 16 do gene *REVI*, c.2614+15A>G (rs717454), estava presente em 9 pacientes com SMD, os quais se encontravam majoritariamente em estágios avançados da doença, de modo que 3 vieram a óbito. Em nossa análise, tal SNP foi considerado como não patogênico, atingindo baixa probabilidade de afetar funções regulatórias segundo o algoritmo utilizado. Tendo em vista esses achados e as referências existentes em publicações anteriores, nas quais a variante em questão foi associada somente a um aumento do risco para distúrbios do comportamento como bipolaridade e depressão, afetando o tratamento farmacológico desses pacientes (XIAO et al., 2017), compreende-se que não há histórico prévio indicativo de sua associação à tumorigênese, e que sua presença em casos de Síndrome Mielodisplásica não possui efeito regulatório funcional aparente.

Atentando também para os SNPs encontrados no gene *POLI*, em estudos prévios Kim e colaboradores realizaram uma análise bioquímica-cinética *in vitro* acerca da consequência da deleção *in frame* p.Asp17delCGA (rs10584411; COSM288348), localizada no éxon 1 do gene *POLI*, na atividade da enzima resultante, e a partir dos resultados obtidos demonstrou-se que tal deleção, a qual não configura mudança no quadro de leitura devido ao número de nucleotídeos removidos ser múltiplo de 3, não causa danos prejudiciais à atuação da enzima mutada Pol  $\tau$ , a qual alcançou desempenho enzimático muito similar à mesma polimerase na sua forma selvagem (KIM et al., 2014). Em nosso estudo, essa variante foi detectada em 10 pacientes, dentre eles, 6 eram de baixo risco de acordo com o IPSS-R, com prognósticos favoráveis e manifestações clínicas leves, referentes aos subtipos SMD-SA-DM (3) e SMD-SA-DU (3). Os demais pacientes pertenciam aos subtipos SMD-EB1 (2) e SMD-EB2 (2), manifestaram sinais clínicos mais graves (> 2 citopenias e >2 displasias) e representavam fenótipos de alto risco e prognósticos desfavoráveis, dos quais 3 vieram a óbito. Tendo em vista sua menção prévia na literatura como uma variante não patogênica que não causa maiores prejuízos à enzima resultante, conjecturamos que esse é um polimorfismo provavelmente comum na população, e que, de forma isolada, não representa um risco significativo para o bom funcionamento do mecanismo TLS.

Em outras duas análises de variantes polimórficas referentes ao gene *POLI*, identificou-se a associação do SNP rs8305 (p.Ala731Thr), localizado no éxon 10, a um risco aumentado para adenocarcinoma pulmonar e carcinoma de células escamosas (XU et al., 2013; SAKIYAMA et al., 2005), e esse mesmo polimorfismo foi mencionado posteriormente em estudos relacionados ao aumento do risco para desenvolvimento de câncer de pulmão, especialmente na presença de fatores etiológicos como exposição à tabaco (MEYERS, 2017). Em nossos pacientes, tal variante foi observada em 4 indivíduos, majoritariamente (3) diagnosticados com SMD em estágio inicial, apresentando quadros clínicos de baixo risco e prognóstico favorável, e segundo análise *in silico* não expressou potencial patogênico em nenhum dos algoritmos empregados (score total 0), e considerando tais achados, infere-se que o rs8305 não está associado ao agravamento dos quadros clínicos em SMD, contudo, devido ao seu histórico de envolvimento no processo da oncogênese, ele poderia estar associado a demais manifestações patogênicas em um contexto de acúmulo de mutações para esse gene.

## 6 CONCLUSÕES

I. As variantes p.Glu571Asp e p.Glu1184\*, referentes ao gene *REVI* e p.Glu122Asp referente ao gene *POLI*, classificadas como patogênicas e detectadas em um alto número de pacientes com SMD, representam boas candidatas a novos marcadores para essa doença, seja em seu estágio inicial ou avançado.

II. O acúmulo de mutações com alto potencial patogênico nos genes *REVI* e *POLI* em pacientes em estágio avançado da SMD, com cariótipo contendo múltiplas alterações citogenéticas e del(7q), está associado a quadros de alto risco e prognóstico desfavorável, denotando o comprometimento da via de tolerância ao dano no DNA e falha do mecanismo de síntese translesão, gerando instabilidade genômica.

III. A baixa frequência de variantes patogênicas nos genes *REVI* e *POLI* identificada em pacientes com SMD e cariótipos contendo nulissomia isolada do cromossomo Y está relacionada ao fenótipo de baixo risco e prognóstico favorável para esses casos, denotando que o mecanismo de síntese translesão não está sofrendo maiores danos nesse contexto.

IV. O perfil genômico misto referente às variantes dos genes *REVI* e *POLI*, identificadas em pacientes com SMD em estágio avançado e cariótipo alterado com múltiplas aneuploidias e del(5q) está relacionado aos quadros de alto risco e prognóstico desfavorável nesses casos, indicando que o mecanismo de síntese translesão pode estar sofrendo ação da instabilidade genômica intrínseca à SMD.

V. A baixa frequência de variantes de modo geral referentes aos genes *REVI* e *POLI* em pacientes com estágio inicial da SMD e cariótipo normal está relacionada ao fenótipo de baixo risco e prognóstico favorável para esses casos, denotando atividade funcional do mecanismo de síntese translesão.

## 7 LIMITAÇÕES E PERSPECTIVAS FUTURAS

A investigação do papel do mecanismo de Síntese Translesão na patogênese da SMD é uma área ainda praticamente inexplorada, de tal modo que, devido à característica ambígua das enzimas DNA polimerases especializadas aqui estudadas, ora como protetoras contra a instabilidade genômica, ora como facilitadoras da oncogênese, estas podem ocupar funções diversas na progressão da doença e na evolução para quadros de leucemia mieloide aguda. Dessa forma, reconhecemos que a presente pesquisa desenvolvida possui limitações, uma vez que não é possível exaurir o escopo de possibilidades existente dentro do tema abordado.

Embora não tenha sido realizada análise estatística que permita traçar coeficientes de associação (ORs) das variantes encontradas aos diferentes fenótipos de risco para SMD, em razão do número diminuto de amostras das quais nós dispomos, o estudo aqui representado faz parte de um estudo guarda-chuva, o qual tem sido desenvolvido já com a perspectiva de aumento do número de amostras sequenciadas, e cuja publicação resultante constará análise estatística correspondente a uma maior robustez de dados.

Também estamos cientes que o sequenciamento de amostras para comparação com linhagens germinativas seria necessário, a fim de determinar quais SNVs e indels são de origem germinativa ou somática, e essa metodologia será incluída em estudos futuros para confirmar a origem das alterações aqui reportadas, assim como também será implementado o aumento do número de amostras sequenciadas e a realização de sequenciamento Sanger para as variantes classificadas como mais importantes em nossa pesquisa. Além disso, planejamos desenvolver também estudos funcionais, com a implementação de linhagens celulares e camundongos *knockout* para conseguir traçar um perfil mais elucidativo sobre a atuação do mecanismo de síntese translesão na Síndrome Mielodisplásica, e, possivelmente, identificar novos biomarcadores prognósticos que possam atuar como peças fundamentais em ambos tratamento e classificação clínica dos nossos pacientes, com o objetivo de traçar novas estratégias que promovam o seu bem-estar e aperfeiçoamento da qualidade de vida.

## 8 REFERÊNCIAS BIBLIOGRÁFICAS

ARBER, D. et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. **Blood**, v. 127, n. 20, p. 2391–2405, 2016.

ASHOUR, M.; MOSAMMAPARAST, N. Mechanisms of damage tolerance and repair during DNA replication. **Nucleic Acids Research**, v. 49, n., p. 3033–3047, 2021.

AWADA, H.; THAPA, B.; VISCONTE, V. The Genomics of Myelodysplastic Syndromes: Origins of Disease Evolution, Biological Pathways, and Prognostic Implications. **Cells**, v. 9, n.11, p. 2512, 2020.

BEJAR, R. et al. Clinical effect of point mutations in myelodysplastic syndromes. **New England Journal of Medicine**, v. 364, n.26, p.2496-506, 2011.

BEJAR, R.; GREENBERG, P.L. The Impact of Somatic and Germline Mutations in Myelodysplastic Syndromes and Related Disorders. **Journal of National Comprehensive Cancer Network.**, v. 15, p. 131–135, 2017.

BEJAR, R.; STEVENSON, K; CAUGHEY, B. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. **Journal of Clinical Oncology**, v. 30, n. 27, p.3376-3382, 2012.

BELLI, C.; PINHEIRO, R.; MAGALHÃES, S. M. et al. Myelodysplastic syndromes in South America: a multinational study of 1080 patients. **American journal of hematology**, v. 90, n. 10, p. 851–858, 2015.

BENNETT, J. et al. Proposals for the classification of the myelodysplastic syndromes. **British journal of haematology**, v. 51, n. 2, p. 189–199, 1982.

BERNARD, E. et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. **Nature medicine**, v. 26, n. 10, p. 1549–1556, 2020.

BERSANELLI, M. et al. Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in Myelodysplastic Syndromes. **Journal of Clinical Oncology**, v. 39, n. 11, p. 1223-1233, 2021.

BOEHM, E.; SPIES, M.; WASHINGTON, M. PCNA tool belts and polymerase bridges form during translesion synthesis. **Nucleic acids research**, v. 44, n. 17, p. 8250–8260, 2016.

BOITEUX, S.; JINKS-ROBERTSON, S. DNA repair mechanisms and the bypass of DNA damage in *Saccharomyces cerevisiae*. **Genetics**, v. 193, n. 102, p. 510-64, 2013.

BOUDSOCQ, F.; LING, H.; YANG, W. Structure-based interpretation of missense mutations in Y-family DNA polymerases and their implications for polymerase function and lesion bypass. **DNA repair**, v. 1, n. 5, p. 343–358, 2002.

BOYLE, A. et al. Annotation of functional variation in personal genomes using RegulomeDB. **Genome Research**, v. 22, n. 9, p.1790-1797, 2012.

BUCKSTEIN, R. Epidemiology, Etiology, and Clinical Presentation of Myelodysplastic Syndromes. *Diagnosis and Management of Myelodysplastic Syndromes*. **Springer, Champ**. p. 3-17, 2020.

CAZZOLA, M.; DELLA PORTA, M.; MALCOVATI, L. The genetic basis of myelodysplasia and its clinical relevance. **Blood**, v. 122, n. 25, p. 4021–4034, 2013.

CECCALDI, R. et al. Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. **Nature**, v. 518, n. 10. p. 258–262, 2015.

CHABANON, R. et al. Targeting the DNA damage response in immuno-oncology: developments and opportunities. **Nature reviews: Cancer**, v. 21, n.11, p. 701–717, 2021.

CHAMSEDDINE, A. et al. Unraveling myelodysplastic syndromes: current knowledge and future directions. **Current oncology reports**, v. 18, n. 1, p. 1-11, 2016.

CHANG D.; CIMPRICH, K. DNA damage tolerance: when it's OK to make mistakes. **Nature Chemical Biology**, v. 5, p. 82–90, 2009.

CHIEREGHIN, C. et al. The Genetics of Myelodysplastic Syndromes: Clinical Relevance. **Genes**, v. 12, n. 8, p. 1144, 2021.

CHOI, Y.; CHAN, A. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. **Bioinformatics**. v. 31, n. 16, p. 2745-2747, 2015.

COSTA, M. et al. Chromosomal abnormalities and dysregulated DNA repair gene expression in farmers exposed to pesticides. **Environmental toxicology and pharmacology**, v. 82, n. 10, p. 564, 2021.

DAIGAKU, Y., DAVIES, A. ; ULRICH, H. Ubiquitin-dependent DNA damage bypass is separable from genome replication. **Nature**, v. 46, n. 5, p. 951–955 (2010).

DE BONT, R.; VAN LAREBEKE, N. Endogenous DNA damage in humans: a review of quantitative data. **Mutagenesis**, v. 19, n. 13, p. 169–185, 2004.

DIDERICH, K.; ALANAZI, M.; HOEIJMAKERS, J. Premature aging and cancer in nucleotide excision repair-disorders. **DNA Repair (Amst)**, v. 10, p. 772–780, 2011.

DOTSON, J.; LEBOWICZ, Y. Myelodysplastic Syndrome. StatPearls. **StatPearls Publishing**, v. 1, p. 10, 2021.

FRANK, E. et al. Translesion replication of benzo[a]pyrene and benzo[c]phenanthrene diol epoxide adducts of deoxyadenosine and deoxyguanosine by human DNA polymerase  $\eta$ . **Nucleic acids research**, v. 30, n.23, p. 5284–5292, 2002.

GARCÍA-ISIDORO, M. et al. Clinical and biological characteristics of myelodysplastic syndromes with nullisomy Y by fish. **Haematologica**, v.82, n. 5, p. 537-41, 1997.

GARCIA-MANERO, G.; CHIEN, K. MONTALBAN-BRAVO, G. Myelodysplastic syndromes: 2021 update on diagnosis, risk stratification and management. **American Journal of Hematology**, v. 95, p. 1399–1420, 2020.

GARCIA-MANERO, G.; CHIEN, K.; MONTALBAN-BRAVO, G. Myelodysplastic syndromes: 2021 update on diagnosis, risk stratification and management. **American Journal of Hematology**, v. 95, p. 1399– 1420, 2020.

GIAGOUNIDIS, A. et al. Biological and Prognostic Significance of Chromosome 5q Deletions in Myeloid Malignancies. **Clinical Cancer Research**, v. 12, n.1, p. 5-10, 2006.

GREENBERG, P. et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. **Blood**, v. 89, n. 6, p. 2079-2088, 1997.

GULILAT, M. et al. Targeted next generation sequencing as a tool for precision medicine. **BMC Medicine Genomics**, v. 12, n. 81, p. 120-124, 2019.

GUO, C. et al. REV1 protein interacts with PCNA: significance of the REV1 BRCT domain in vitro and in vivo. **Molecular cell**, v. 23, n.2, p. 265–271, 2006

HE, X. et al., REV1 genetic variants associated with the risk of cervical carcinoma. **European journal of epidemiology**, v. 23, n. 6, p. 403–409, 2008.

HELLSTRÖM-LINDBERG, E.; TOBIASSON, M.; GREENBERG, P. Myelodysplastic syndromes: moving towards personalized management. **Haematologica**, v. 105, n. 7, p. 1765–1779, 2020.

JAIN, R.; AGGARWAL, A.; RECHKOBLIT, O. Eukaryotic DNA polymerases. **Current opinion in structural biology**, v. 53, p. 77–87, 2018.

KANEKO, H. et al. TP53 Mutations Emerge at Early Phase of Myelodysplastic Syndrome and Are Associated With Complex Chromosomal Abnormalities. **Blood**, v.85, n. 8, p. 2189-2193, 1995.

KANTARJIAN, H. ; O'BRIEN, S. ; RAVANDI, F. Proposal for a new risk model in myelodysplastic syndrome that accounts for events not considered in the original International Prognostic Scoring System. **Cancer**, v. 113, n. 6, p. 1351-1361, 2008

KENNEDY, A.; SHIMAMURA, A. Genetic predisposition to MDS: clinical features and clonal evolution. **Blood**, v. 133, n.10, 1071–1085, 2019.

KULASEKARARAJ, A. et al. TP53 mutations in myelodysplastic syndrome are strongly correlated with aberrations of chromosome 5, and correlate with adverse prognosis. **British journal of haematology**, v. 160, n. 5, p. 660–672, 2013.

LANGIE, S. A. et al. Causes of genome instability: the effect of low dose chemical exposures in modern society. **Carcinogenesis**, v. 36, n. 1, p. 61-88, 2015.

LINDAHL, T. Instability and decay of the primary structure of DNA. **Nature**, v. 362, n. 6, p. 709-15, 1993.

LODÉ, L. et al. Emergence and evolution of TP53 mutations are key features of disease progression in myelodysplastic patients with lower-risk del(5q) treated with lenalidomide. **Haematologica**, v. 103, n. 4, p. 143-146, 2017.

MAEDA, T. Treatment for myelodysplastic syndromes and future perspectives in the genomic era. **The Japanese journal of clinical hematology**, v. 62, n. 8, p. 1038–1049, 2021.

MAGALHÃES, S. et al. Guidelines on myelodysplastic syndromes: Associação Brasileira de Hematologia, Hemoterapia e Terapia Celular. **Hematology, Transfusion and Cell Therapy**., v. 40, n. 3, p. 255-261, 2018.

MAGALHÃES, S. M. et al. Epidemiological and Clinicopathological Data From the Brazilian Registry of Patients with Myelodysplastic Syndromes and Comparative Analysis Between Different Geographic Areas. **Blood**, v. 116, n. 21, p. 1884, 2010.

MAKRIDAKIS, N.; REICHARDT, J. Translesion DNA polymerases and cancer. **Frontiers in genetics**, v. 3, p. 174, 2012.

MALCOVATI, L.; CAZZOLA, M. et al. SF3B1-mutant MDS as a distinct disease subtype: a proposal from the International Working Group for the Prognosis of MDS. **Blood**, v. 136, n. 2, p. 157–170, 2020.

MARTINS, C.; PUTON, C.; SOUSA, L. E.; SILVA, A. M. Myelodysplastic syndrome mortality profile from 2014 to 2018. **Acta Scientiae Anatomica**, v. 1, n. 2, p. 111-112, 2021.

MASUTANI, C et al. The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase  $\epsilon$ . **Nature**, v. 399, n. 6737, p. 700–704, 1999.

MAUL, R. et al. DNA polymerase  $\epsilon$  functions in the generation of tandem mutations during somatic hypermutation of antibody  $\epsilon$  genes. **The Journal of experimental medicine**, v. 213, n. 9, p. 1675–1683, 2016.

MCGOWAN-JORDAN, J. ISCN 2016: An international system for human cytogenomic nomenclature (2016); recommendations of the international standing human committee on human cytogenomic nomenclature including new sequence-based cytogenomic. **Karger**, 2016.

NAKAJIMA, H. Molecular Pathogenesis and Treatment of Myelodysplastic Syndromes. **Internal medicine (Tokyo, Japan)**, v. 60, n. 1, p. 15–23, 2021.

OLIVEIRA, R. et al. c.9253-6T > c REV3L: A novel marker of poor prognosis in Myelodysplastic syndrome. **Hematology, Transfusion and Cell Therapy**, v. 43, n. 03, 2021.

OUSEPH, M.; HASSERJIAN, R. et al. Genomic alterations in patients with somatic loss of the Y chromosome as the sole cytogenetic finding in bone marrow cells. **Haematologica**, v. 106, n. 2, p. 555–564, 2021.

PEJAVER, V. et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. **Nature Communications**, v. 11, n. 5918, 2020.

PERAZZIO, A.; CHAUFFAILLE, M. When karyotype is decisive for myelodysplastic syndromes diagnosis. **Hematology, transfusion and cell therapy**, v. 41, p. 371-373, 2019.

PINHEIRO, R. F.; CHAUFFAILLE, M. Comparison of I-FISH and G-banding for the detection of chromosomal abnormalities during the evolution of myelodysplastic syndrome. **Brazilian Journal of Medical and Biological Research**, v. 42, n. 11, p. 1110-1112, 2009.

POUSADA, G. et al. Mutational and clinical analysis of the ENG gene in patients with pulmonary arterial hypertension. **BMC Genetics**, v.17, n. 1, p. 7–9, 2016.

Powers, K. T., & Washington, M. T. (2018). Eukaryotic translesion synthesis: Choosing the right tool for the job. **DNA repair**, v. 71, p. 127–134.

QIN, D. Next-generation sequencing and its clinical application. **Cancer biology & medicine**, v. 16, n. 1, p. 4–10, 2019.

QUINET, A., LERNER, L. K., MARTINS, D. J. Filling gaps in translesion DNA synthesis in human cells. **Mutation research. Genetic toxicology and environmental mutagenesis**, v. 836, p. 127–142, 2018.

RAZA, A.; IVERSON, N.; ALI, A. Developments in the treatment of transfusion-dependent anemia in patients with myelodysplastic syndromes: epidemiology, etiology, genetics, and targeted therapies. **Advances in Genomics and Genetics**, v. 4, p. 95, 2014.

REILLY, A. et al. Deletions of 5q Promote Genome Instability in TP53-Mutant MDS/AML. **Blood**, v. 134, n. 12, p. 1244-1247, 2019.

RIBEIRO, H. L. et al. ATM polymorphism is associated with low risk myelodysplastic syndrome. **DNA repair**, v. 12, n. 2, p. 87–89, 2013.

RIBEIRO, H. L. et al. Polymorphisms of DNA repair genes are related to the pathogenesis of myelodysplastic syndrome. **Hematological oncology**, v. 33, n. 4, p. 220–228, 2015.

ROBIN, M., DE WITTE, T. Myelodysplastic Syndromes. The EBMT Handbook: Hematopoietic Stem Cell Transplantation and Cellular Therapies. **Springer Cham (CH)**, v. 7, c. 73, 2019.

SAKIYAMA, T. et al. Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. **International journal of cancer**, v. 114, n. 5, 730–737, 2005.

SAYGIN, C.; CARRAWAY, H. Current and emerging strategies for management of myelodysplastic syndromes. **Blood reviews**, v. 48, n. 10, p. 791, 2021.

SCHANZ J. ; TUCHLER H. ; SOLE, F. New comprehensive cytogenetic scoring system for primary myelodysplastic syndromes (MDS) and oligoblastic acute myeloid leukemia after MDS derived from an international database merge. **Journal of Clinical Oncology**, v. 30, n. 2, p. 820–829, 2012.

SCHARENBERG, C. et al. Progression in patients with low- and intermediate-1-risk del(5q) myelodysplastic syndromes is predicted by a limited subset of mutations. **Haematologica**, v. 102, n.3, p. 503-507, 2017.

SCHWARZ, J. et al. MutationTaster2: mutation prediction for the deep-sequencing age. **Nature Methods**, v. 11, n. 4, p. 361-2, 2014.

SHAHRABI, S. et al. Sex chromosome changes in leukemia: cytogenetics and molecular aspects. **Hematology (Amsterdam, Netherlands)**, v. 23, n. 3, p. 139–147, 2018.

SHARMA, S.; HELCHOWSKI, C.; CANMAN, C. The roles of DNA polymerase  $\zeta$  and the Y family DNA polymerases in promoting or preventing genome instability. **Mutations Research**, v. 97, n.110, p. 743-744, 2013.

SIM, N. et al. SIFT web server: predicting effects of amino acid substitutions on proteins. **Nucleic Acids Research**, v. 40, p. 542-7, 2012.

SPAULDING, T.; STOCKTON, S.; SAVONA, M.. The evolving role of next generation sequencing in myelodysplastic syndromes. **British Journal of Haematology**, v. 188, p. 224-239, 2020.

STEENSMA, D. P. The evolving role of genomic testing in assessing prognosis of patients with myelodysplastic syndromes. **Best practice & research. Clinical haematology**, v. 30, n. 4, p. 295–300, 2017.

TISSIER, A.; KANNOUCHE, P.; RECK, M. P.Co-localization in replication foci and interaction of human Y-family members, DNA polymerase pol eta and REV1 protein. **DNA repair**, v. 3, n.11, p. 1503–1514, 2004.

TONZI, P. et al. Role of Y-family translesion DNA polymerases in replication stress: Implications for new cancer therapeutic targets. **DNA Repair**, v. 78, n. 3, p. 20-26, 2019.

VAISMAN, A., & WOODGATE, R. (2020). Mysterious and fascinating: DNA polymerase  $\iota$  remains enigmatic 20 years after its discovery. **DNA repair**, v. 93, p. 102-107.

VARDIMAN, J.; HARRIS, N.; BRUNNING, R. The World Health Organization (WHO) classification of the myeloid neoplasms. **Blood**, v. 100, n.7, p. 2292–302, 2002.

VAZ-DRAGO, R.; CUSTÓDIO, N.; CARMO-FONSECA, M. Deep intronic mutations and human disease. **Human Genetics**, v. 136, n. 10, p. 1093–1111, 2017.

VAZIRI, C., et al. Unravelling roles of error-prone DNA polymerases in shaping cancer genomes. **Oncogene**, v. 40, p. 6549–6565, 2021.

WANG, R. et al. DNA polymerase  $\iota$  compensates for Fanconi anemia pathway deficiency by countering DNA replication stress. **Proceedings of the National Academy of Sciences of the United States of America**, v. 117, n. 52, p. 33436–33445, 2020.

WANG, Y. et al. Evidence that in xeroderma pigmentosum variant cells, which lack DNA polymerase eta, DNA polymerase iota causes the very high frequency and unique spectrum of UV-induced mutations. **Cancer research**, v. 67, n. 7, p. 3018–3026, 2007.

WILSON, D.; BOHR, V. The mechanics of base excision repair, and its relationship to aging and disease. **DNA Repair (Amst)**, v. 6, p. 544–559, 2007.

XU, Y. et al. Implications of mutational spectrum in myelodysplastic syndromes based on targeted next-generation sequencing. **Oncotarget**, v. 8, n. 47, p. 824–829, 2017.

YEOM, M. et al. Three Human Pol  $\tau$  Variants with Impaired Polymerase Activity Fail to Rescue H<sub>2</sub>O<sub>2</sub> Sensitivity in POLI-Deficient Cells. **Chemical Research in Toxicology**, v. 33, n. 8, p. 2120-2129, 2020.

YOUSEFZADEH, M. ; HENPITA, C.;VYAS, R. DNA damage—how and why we age?. **eLife**, v. 10, p. 62, 2021.

ZHANG, L. et al. Molecular genetic evidence of Y chromosome loss in male patients with hematological disorders. **Chinese medical journal**, v. 120, n. 22, p. 2002–2005, 2007.

ZHOU, B.; PATA, J.; STEITZ, T. Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. **Molecular cell**, v. 8, n. 2, p. 427–437, 2001.

## APÊNDICES

### APÊNDICE A: Termo de Consentimento Livre e Esclarecido



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**FACULDADE DE MEDICINA**  
**DEPARTAMENTO DE MEDICINA CLÍNICA**

#### **TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**

**PROJETO:** ANÁLISE DE EXPRESSÃO GÊNICA E MUTAÇÕES DAS ENZIMAS DNA POLIMERASES COM ATIVIDADE TLS EM PACIENTES PORTADORES DE SÍNDROME MIELODISPLÁSICA

Você está sendo convidado a participar de um projeto de pesquisa. Sua participação é importante, porém, você não deve participar contra a sua vontade. Leia atentamente as informações abaixo e faça qualquer pergunta que desejar.

O abaixo assinado,

\_\_\_\_\_, \_\_\_\_ anos, RG nº

\_\_\_\_\_, declara que é de livre e espontânea vontade que está participando como voluntário do projeto de pesquisa supracitado, de responsabilidade do pesquisador Roberta Taiane Germano de Oliveira / Prof. Dr. Ronald Feitosa Pinheiro. O abaixo assinado está ciente de que:

#### **NATUREZA E PROPÓSITO DO ESTUDO**

O objetivo da pesquisa é estudar o material genético de participantes portadores de Síndrome Mielodisplásica gerando dados que favoreçam a uma melhor compreensão do

surgimento e evolução clínica desta doença.

## **PROCEDIMENTOS A SEREM REALIZADOS E RESPONSABILIDADES**

A amostra biológica utilizada na presente pesquisa corresponde a medula óssea do participante. A medula óssea corresponde a um tecido líquido-gelatinoso que ocupa o interior dos ossos, sendo conhecida popularmente por 'tutano'. Na medula óssea são produzidos os componentes do sangue: as hemácias (glóbulos vermelhos), os leucócitos (glóbulos brancos) e as plaquetas. A coleta da medula óssea será realizada por médico hematologista experiente com agulha de mielograma mediante punção esternal. O osso do esterno é um osso chato, plano e ímpar localizado no centro do tórax. O osso do esterno é um importante osso hematopoético, ou seja, de produção das células sanguíneas. Serão coletadas somente 3mL de amostra de medula óssea com o uso de uma agulha específica para aspiração da medula óssea. Todo o procedimento de coleta da medula óssea é realizado mediante administração de anestésico local com duração máxima de 15 minutos. O procedimento de coleta esternal da medula óssea, pode, raramente, determinar uma equimose (mancha arroxeadas) ao redor do local de onde foi retirado, desaparecendo em poucos dias, e poderá ocorrer dor discreta e de fácil alívio, podendo ser, ocasionalmente, dor de maior intensidade. Excepcionalmente, poderá ocorrer sangramento local. Raramente, pode ocorrer infecção local. Para participantes com maior sensibilidade dolorosa existe a possibilidade de realizar o procedimento sob anestesia geral. A recoleta deste material é necessária, em poucos casos, pela amostra ser insuficiente ou inadequada para análise. Após o procedimento, serão coletados dados pessoais de sexo e idade, apresentação clínica e checagem dos exames laboratoriais indicados para esclarecimento diagnóstico, tais como: hemograma completo, citologia e histologia (análise microscópica das células) da medula óssea, avaliação para depósitos de ferro medular, bem como outros exames que são feitos mesmo para os participantes que não participam de nenhuma pesquisa. É de sua responsabilidade: comparecer nas datas e horários informados e submeter-se aos procedimentos de rotina do serviço.

## **PARTICIPAÇÃO VOLUNTÁRIA**

Sua participação é voluntária, você não receberá nenhum ressarcimento monetário por participar da pesquisa e tem a liberdade de desistir ou interromper a participação neste estudo no momento em que desejar. Neste caso, você deve informar imediatamente sua decisão ao

pesquisador responsável ou a qualquer um membro de sua equipe, sem necessidade de qualquer explicação e sem que isto venha interferir no seu atendimento nesta instituição. Independentemente de seu desejo e consentimento, sua participação no estudo poderá ser interrompida, em função da ocorrência de qualquer doença que, a critério médico, prejudique a continuação de sua participação no estudo, do não cumprimento das normas estabelecidas, de qualquer outro motivo que, a critério da pesquisadora, seja do interesse de seu próprio bem-estar ou dos demais participantes e, por fim, da suspensão do estudo como um todo. O Laboratório de Citogenômica do Câncer o manterá informado, em tempo oportuno, sempre que houver alguma informação adicional que possa influenciar seu desejo de continuar participando no estudo e prestará qualquer tipo de esclarecimento em relação ao progresso da pesquisa, conforme sua solicitação.

### **DIVULGAÇÃO DE INFORMAÇÕES QUANTO À PARTICIPAÇÃO NO ESTUDO**

Os registros que possam identificar sua identidade serão mantidos em sigilo, a não ser que haja obrigação legal de divulgação. Você não será identificado por ocasião da publicação dos resultados obtidos. Contudo, o(s) monitor(es) do Estudo, auditor(es), membros do Comitê de Ética em Pesquisa com Seres Humanos, ou autoridades do(s) órgão(s) regulamentar(es) envolvido(s) terão direito de ter acesso aos registros originais de dados clínicos de sua pessoa, coletados durante a pesquisa, na extensão em que for permitido pela Lei e regulamentações aplicáveis, como o propósito de verificar os procedimentos e dados do estudo, sem, no entanto, violar a condição de que tais informações são confidenciais.

### **CONTATOS E PERGUNTAS**

Caso surja algum imprevisto ou dúvidas, você deverá entrar em contato solicitar contato direto com o pesquisador responsável pelo estudo: Roberta Taiane Germano de Oliveira (99662327) ou com seu Orientador acadêmico: Prof. Dr. Ronald Feitosa Pinheiro (81881972) ou entre em contato com o Comitê de Ética em Pesquisa da UFC/PROPESQ - Rua Coronel Nunes de Melo, 1000 - Rodolfo Teófilo / Telefone: 3366- 8344/46. (Horário: 08:00-12:00 horas de segunda a sexta-feira). O CEP/UFC/PROPESQ é a instância da Universidade Federal do Ceará responsável pela avaliação e acompanhamento dos aspectos éticos de todas as pesquisas envolvendo seres humanos. Somente assine este termo se você tiver a certeza de que recebeu todos os esclarecimentos e informações para decidir conscientemente sobre a sua participação neste estudo.

**ASSINATURAS**

Autorizo o acesso às minhas informações de saúde aos membros da equipe de pesquisadores, nas condições estabelecidas descritas nos itens acima. Não renunciei qualquer direito legal que eu venha a ter ao participar deste Estudo. Não receberei nenhum ressarcimento monetário por participar da pesquisa. Eu, por fim, declaro que li cuidadosamente todo este documento denominado Termo de Consentimento Livre e Esclarecido e que, após a assinatura, tive oportunidade de fazer perguntas sobre o conteúdo do mesmo e também sobre o referido estudo, recebendo explicações que responderam por completo minhas dúvidas e reafirmando estar livre e espontaneamente decidido a participar do estudo, ficando munido de uma via do documento assinado pelo pesquisador responsável.

\_\_\_\_/\_\_\_\_/\_\_\_\_

Data

\_\_\_\_\_

Assinatura do participante da pesquisa

\_\_\_\_/\_\_\_\_/\_\_\_\_

Data

\_\_\_\_\_

Assinatura do pesquisador responsável

\_\_\_\_/\_\_\_\_/\_\_\_\_

Data

\_\_\_\_\_

Assinatura do responsável pela aplicação do TCLE

## APÊNDICE B: CARACTERIZAÇÃO DOS PACIENTES EM RELAÇÃO ÀS VARIÁVEIS CLÍNICAS E SOCIODEMOGRÁFICAS

No presente estudo, foi realizado o sequenciamento alvo dos genes *REVI* e *POLI* a partir das amostras de medula óssea de 14 pacientes, sendo destes 7 (50%) indivíduos do sexo masculino e 7 (50%) indivíduos do sexo feminino (Tabela 12).

Com relação à idade dos indivíduos, a média foi de 70,6 anos e a mediana 72,5 anos (mínimo de 28 anos e máximo de 91 anos). A variável idade dividida em dois grupos: idade categorizada ( $\leq 60$  anos e  $> 60$  anos) e idade categorizada de acordo com valores descritos no IPSS-R ( $\leq 60$  anos,  $> 60 - \leq 70$  anos,  $> 70 - \leq 80$  anos e  $> 80$  anos) (GREENBERG et al., 2012). Na primeira categorização, os indivíduos apresentaram maior concentração no grupo acima de 60 anos (85,7%), contra os 14,2% restantes pertencentes ao grupo abaixo de 60 anos. Em relação à segunda divisão, a distribuição foi mais homogênea, sendo 14,2% pertencente ao grupo  $\leq 60$  anos, 28,5% no grupo  $> 60 - \leq 70$  anos, 28,5% no grupo  $> 70 - \leq 80$  anos e 28,5% no grupo acima de 80 anos (Tabela 12).

Quanto à origem dos pacientes, identificaram-se concentrações relativamente semelhantes entre indivíduos oriundos da zona urbana (57,1%) e pertencentes à zona rural (42,8%). Porcentagens similares foram obtidas em relação ao histórico de exposição dos pacientes a agentes tóxicos, sendo 57,1% dos casos associados a exposição contínua a componentes com alto potencial carcinogênico, dentre eles agrotóxicos, inseticidas industriais, solventes e diluentes de tintas, etílicos, querosene, tabaco e óleo diesel, enquanto 42,8% não possuíam histórico de exposição (Tabela 12)

Em relação à classificação clínica dos subgrupos de SMD identificados no diagnóstico dos pacientes, todos foram classificados como SMD primária, e seus respectivos subgrupos foram determinados de acordo com a revisão mais recente da classificação de neoplasias hematológicas da OMS (ARBER et al., 2016). Verificou-se predomínio de formas iniciais da doença (64, 2%), sendo 4 indivíduos (28,5%) portadores de SMD-SA-DU e 5 (35,7%) portadores de SMD-SA-DM. Em contrapartida, 35,7% dos pacientes foram diagnosticados com formas mais graves da doença, dentre eles 2 (14,2%) com SMD-EB 1 e 3 (21,4%) com SMD EB 2. Todas as informações estão dispostas na tabela 12.

Tabela 12 – Caracterização das variáveis sociodemográficas e classificação clínica dos pacientes com SMD.

Variáveis		Nº	%
Sexo	Feminino	7	50%
	Masculino	7	50%
Idade (categorizada)	≤ 60 anos	2	14,2%
	> 60 anos	12	85,7%
Idade (Categorizada de acordo com o IPSS 2012)	≤60 anos	2	14,2%
	>60 - ≤70 anos	4	28,5%
	>70 - ≤80 anos	4	28,5%
	>80 anos	4	28,5%
Origem	Urbana	8	57,1%
	Rural	6	42,8%
Exposição a tóxicos	Sim	8	57,1%
	Não	6	42,8%
Etiologia	SMD Primária	14	100%
	SMD Secundária	0	0%
Classificação da OMS (2016)	SMD-SA-DU	4	28,5%
	SMD-SA-DM	5	35,7%
	SMD-EB 1	2	14,2%
	SMD-EB 2	3	21,4%
Formas da SMD	Iniciais	9	64,2%
	Avançadas	5	35,7%

Fonte: Elaborada pela autora

No tocante às variáveis clínicas dos pacientes no presente estudo, 11 pacientes apresentaram biópsia medular. Dentre eles, 1 paciente apresentou medula hipocelular (7,4%) e 10 pacientes apresentavam medula hipercelular (92,8%), não houve pacientes apresentando medula normocelular. Em referência aos dados obtidos através do mielograma, predominaram os pacientes com displasia em pelo menos duas linhagens (50%), seguidos dos pacientes com displasias em 3 linhagens (21,4%) e por fim os pacientes com displasia em apenas uma linhagem (28,5%). Todos os 14 pacientes apresentaram diseritropoese, enquanto

8 pacientes apresentaram dismegacariopese (57,1%) e 9 apresentaram disgranulopoiese (64,2%).

Para a porcentagem de blastos, maioria dos pacientes demonstraram possuir  $\leq 2\%$  de blastos (64,2%), enquanto 3 apresentaram faixa de porcentagem de blastos  $>5\% - \leq 10\%$  e 2 pacientes com faixa de blastos acima de 10%. Em relação às porcentagens de sideroblastos em anel, maioria (64,2%) dos pacientes apresentou porcentagem acima de 15%, 1 paciente apresentou faixa de sideroblastos em anel entre  $\geq 1\% - \leq 14\%$  e os demais não apresentaram sideroblastos em anel.

Quanto às variáveis relativas a achados de sangue periférico, identificou-se a seguinte distribuição para o número de citopenias encontradas: 1 paciente não apresentou citopenias no sangue periférico (7,1%), 6 pacientes apresentaram 1 citopenia (42,8%), 3 pacientes apresentavam 2 citopenias (21,4%) e por fim 4 pacientes apresentavam 3 citopenias (28,5%). A determinação dos níveis de hemoglobina no sangue revelou que maioria (50%) dos pacientes possuía níveis abaixo de 8g/dL, em comparação a 28,5% dos pacientes que apresentaram faixa entre 8 - <10 e 21,4% dos pacientes que apresentaram nível acima de 10g/dL. Para a contagem absoluta de neutrófilos (ANC), observou-se que 9 pacientes possuíam níveis acima ou iguais a 800/mm<sup>3</sup>, enquanto 5 apresentaram contagem abaixo de 800/mm<sup>3</sup>. A aferição da concentração de plaquetas no sangue revelou que grande parte dos pacientes (71,4%) possuía níveis acima ou iguais a 100/mm<sup>3</sup>, enquanto os demais (28,5%) apresentaram níveis abaixo de 50/mm<sup>3</sup>. As referidas informações estão compiladas na Tabela 13 abaixo.

Tabela 13 - Caracterização descritiva das variáveis clínicas dos pacientes com SMD.

Variáveis		Nº	%
Celularidade da medula óssea	Hipocelular	1	7,1%
	Normocelular	0	0%
	Hiper celular	10	92,8%
Displasias na medula óssea	1	4	28,5%
	2	7	50%
	3	3	21,4%
Diseritropoese	Presença	14	100%
	Ausência	0	0%

Dismegacariopose	Presença	8	57,1%
	Ausência	6	42,8%
Disgranulopose	Presença	5	35,7%
	Ausência	9	64,2%
Faixa das percentagens de Sideroblastos em anel	0%	4	28,5%
	≥1% - ≤14%	1	7,1%
	≥15%	9	64,2%
Faixas das percentagens de blastos	≤2%	9	64,2%
	>2 a <5%	0	0%
	>5% - ≤10%	3	21,4%
	> 10%	2	14,2%
Número de citopenias no sangue periférico	0	1	7,1%
	1	6	42,8%
	2	3	21,4%
	3	4	28,5%
Hemoglobina (g/dL)	≥10	3	21,4%
	8 - <10	4	28,5%
	<8	7	50%
Contagem absoluta de neutrófilos (ANC) (mm <sup>3</sup> )	≥800	9	64,2%
	<800	5	35,7%
Plaquetas (mm <sup>3</sup> )	≥100	10	71,4%
	50 - <100	0	0%
	<50	4	28,5%

Fonte: Elaborada pela autora.

A análise citogenética realizada através da cariotipagem dos pacientes permitiu determinar a frequência das alterações nas amostras medulares. Dos 14 pacientes, 4 apresentaram ausência de metáfase. Dos 10 pacientes restantes, identificou-se que 5 indivíduos possuíam cariótipo normal, em contraste à outra metade dos pacientes que possuía cariótipo alterado, sendo estes últimos com presença de aneuploidia. Dentre as alterações

encontradas comumente associadas à SMD, 1 paciente apresentou deleção no cromossomo 5 - del(5q) e 1 paciente apresentou deleção no cromossomo 7 - del(7q).

Em relação à estratificação prognóstica, com os pacientes em que foi possível aplicar parâmetros do IPSS-R para tal análise, 5 apresentaram prognóstico favorável (35,7%), 2 apresentaram prognóstico muito favorável (14,5%), enquanto os demais apresentaram prognóstico intermediário (7,1%), desfavorável (14,2%) ou muito desfavorável (7,1%). Quanto ao grupo de risco determinado através IPSS-R, 4 pacientes pertenciam ao grupo de risco muito baixo (28,5%), enquanto 3 pacientes pertenciam ao grupo de alto risco (21,4%) (Tabela 14).

Tabela 14 – Frequência dos resultados citogenéticos e seu impacto prognóstico em pacientes com SMD ao diagnóstico.

<b>Variável</b>		<b>N</b>	<b>%</b>
Metáfase	Presença	10	71,4%
	Ausência	4	28,5%
Cariótipo	Normal	5	35,7%
	Alterado	5	35,7%
Aneuploidia	Alterado com Aneuploidia	5	35,7%
	Alterado sem Aneuploidia	0	0%
Alterações no cromossomo 5	Alterado com 5q-	1	7,1%
	Alterado sem 5q-	4	28,5%
Alterações no cromossomo 7	Alterado com 7q-	1	7,1%
	Alterado sem 7q-	4	28,5%
Prognóstico (IPSS-R)	Muito favorável	2	14,2%
	Favorável	5	35,7%
	Intermediário	1	7,1%
	Desfavorável	2	14,2%
	Muito desfavorável	1	7,1%
Grupo de risco (IPSS-R)	Risco muito baixo	4	28,5%
	Baixo risco	1	7,1%
	Intermediário	0	0%
	Alto risco	3	21,4%
	Risco muito alto	1	7,1%

Fonte: Elaborado pela autora.

No tocante ao seguimento terapêutico dos pacientes durante seu histórico em SMD, observou-se que metade dos indivíduos tornaram-se dependentes transfusionais, adotando o critério de 1 transfusão a cada 8 semanas durante o período de 4 meses estabelecido por Malcovati et al. (2005). Além disso, durante o decorrer da execução do presente estudo, nenhum paciente evoluiu para quadro de Leucemia Mieloide Aguda (LMA), no entanto, 3 indivíduos vieram a óbito. A descrição clínica detalhada dos 14 pacientes incluídos no estudo está disposta na tabela 15.

Tabela 15 – Descrição clínica dos pacientes com SMD incluídos no estudo.

Paciente	Sexo	Idade	Celularidade	Cariótipo – ISCN 2016	OMS 2016	IPSS-R	DT	e-LMA	Óbito
1	F	28	-	46 XX [20]	SMD-SA-DU	Risco muito baixo	Sim	Não	Não
2	M	79	Hipocelular	46,XY,del(17)(p11.2)[3]/47,XY,+mar[3]/47,XY,+20[4]/46,XY,del(17)(p11.2),+mar[2]/46,XY[9]	SMD-EB 1	Alto risco	Não	Não	Sim
3	F	70	Hiperclular	Ausência de metáfase	SMD-SA-DM	-	Sim	Não	Não
4	M	63	Hiperclular	47,XY,+8[6]/47,XY,del(7)(q32),+8[7]/46,XY[2]	SMD -EB 2	Alto risco	Sim	Não	Sim
5	M	91	-	46,X,-Y[4]/46,XY[16]	SMD-SA-DU	Risco muito baixo	Não	Não	Não
6	F	51	Hiperclular	46,XX[11]	SMD-SA-DU	Risco muito baixo	Não	Não	Não
7	M	67	Hiperclular	46,XY[20]	SMD-EB 1	Alto risco	Não	Não	Não
8	M	63	Hiperclular	37,X,-2,-3,-9,-11,-12,-15,-16,-18,-Y[8]/46,XY,del(5)(q15q33)[5]/46,XY[6]	SMD-EB 2	Risco muito alto	Sim	Não	Não
9	F	72	Hiperclular	46, XX [25]	SMD-SA-DM	Baixo risco	Sim	Não	-
10	F	84	Hiperclular	Ausência de metáfase	SMD-SA-DU	-	Não	Não	Não
11	F	87	Hiperclular	Ausência de metáfase	SMD-SA-DM	-	Não	Não	Não
12	F	81	Hiperclular	46,XX[5]	SMD-SA-DM	Risco muito baixo	Sim	Não	Não
13	M	80	Hiperclular	45,X,-Y[18]/46,XY[7]	SMD-SA-DM	Risco muito baixo	Não	Não	Não
14	M	73	-	Ausência de metáfase	SMD-EB 2	-	Sim	Não	Sim

DT: Dependência transfusional, e-LMA: evolução para Leucemia Mieloide Aguda

Fonte: Elaborada pela autora