



UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO



ANDRÉ LUIZ PINTO SANTOS

**Diagnóstico genético, gestão e visualização de dados genômicos de pacientes
com síndromes de falência medular hereditária**

Ribeirão Preto - SP

2023

ANDRÉ LUIZ PINTO SANTOS

**Diagnóstico genético, gestão e visualização de dados genômicos de pacientes
com síndromes de falência medular hereditária**

Versão original

Tese apresentada ao Programa de pós-graduação em Oncologia Clínica, Células-tronco e Terapia Celular da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP) para obtenção do título de Doutor em Ciências.

Área de concentração: Diferenciação Celular Normal e Neoplásica.

Orientador: Prof. Dr Rodrigo do Tocantins Calado De Saloma Rodrigues

Ribeirão Preto - SP

2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA

PINTO, André Luiz

Diagnóstico genético, gestão e visualização de dados genômicos de pacientes com síndromes de falência medular hereditária/ André Luiz Pinto Santos; orientador, Rodrigo do Tocantins Calado. Ribeirão Preto : FMRP, 2023.

109 p.: il. ; 30 cm

Tese (Doutorado em Ciências) - Programa de pós-graduação em Oncologia Clínica, Células-tronco e Terapia Celular, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 2023

Versão original

1. Síndromes de falência medular hereditária. 2. Diagnóstico genético. 3. Gestão e visualização de dados.

Nome: PINTO, André Luiz

Título: Diagnóstico genético, gestão e visualização de dados genômicos de pacientes com síndromes de falência medular hereditária

Tese apresentada ao Programa de pós-graduação em Oncologia Clínica, Células-tronco e Terapia Celular da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP) para obtenção do título de Doutor em Ciências. Área de concentração: Diferenciação Celular Normal e Neoplásica.

Orientador: Prof. Dr Rodrigo do Tocantins Calado

Aprovado em: ___/___/_____

Banca examinadora

Orientador: Prof. Dr.: _____

Instituição: _____ Assinatura: _____

Prof(a). Dr(a).: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof(a). Dr(a).: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof(a). Dr(a).: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

APOIO E SUPORTE FINANCEIRO

Este trabalho foi realizado com o apoio financeiro das seguintes entidades e instituições:

- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)
 - Código de Financiamento 001
- Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)
 - Centro de Terapia Celular – CTC (CEPID)
- Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
 - INCT em Células-Tronco e Terapia Celular no Câncer
 - Bolsa DTI

Período: 14/12/2018 a 31/01/2019
- Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP)
- Fundação Hemocentro de Ribeirão Preto
- Fundação de Apoio ao Ensino, Pesquisa e Assistência do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FAEPA-HC-FMRP-USP)

AGRADECIMENTOS

A Deus, o criador de todas as leis naturais e sobrenaturais, a quem pude conhecer melhor em meio a tantas possibilidades que a academia/razão nos oferece.

Aos meus pais, Emanuel e Elizabete, as pessoas mais importantes da minha vida, fonte primordial de tudo que sou, de tudo que tenho e do que ainda conquistarei. Eles me ensinaram a responsabilidade que o mundo me cobra diariamente e a honestidade dos homens bons.

Aos meus queridos irmãos, Vinícius e Manu, fiéis incentivadores da minha caminhada e fortalezas nas minhas ausências. Que sigamos sempre unidos para retribuirmos juntos tudo o que nossos pais heróis nos proporcionaram.

Ao meu orientador, Professor Rodrigo Calado, a quem devo a parte mais importante da minha formação científica. Obrigado por fortalecer meu senso crítico, por me desafiar sempre e confiar em mim quando foi preciso.

Aos pacientes participantes desse e de outros estudos com doenças raras. Agradeço por colaborar e confiar no trabalho da Ciência. Espero que nós cientistas possamos retribuir, elucidando as causas e encontrando tratamentos para reduzir o sofrimento de milhões.

À Flávia Donaires, fiel companheira nas labutas diárias e pessoa indispensável nesse e em todos os trabalhos acadêmicos nos últimos 8 anos. Obrigado também pela amizade e por falar tantas bobagens tornando nossos dias menos difíceis. Saiba que tenho grande admiração por você.

À Bárbara e à Ritinha, sustentáculos do laboratório, cuidando das amostras, dos insumos e do ambiente pré-analítico para que nosso trabalho seja mais produtivo. Além disso, obrigado pela amizade.

Ao Diego Clé, ao Luiz Fernando, aos residentes e a todo corpo clínico do Laboratório Transnacional de Hematologia. Agradeço pelo compartilhamento de conhecimento e por levar nosso trabalho da bancada aos ambulatórios.

Fernanda Borges, Maitê, Sabrina e Aline. Obrigado pela amizade e pela companhia nesses anos de vida acadêmica/profissional.

Ao Sabin Diagnóstico e Saúde, ao Hemocentro de Ribeirão Preto e ao Hospital de Amor. Agradeço as essas empresas que acreditaram no meu potencial, deram-me oportunidades de me tornar um profissional, mostraram-me as necessidades da sociedade e me ensinaram pragmatismo do mundo corporativo.

A CAPES, ao CNPq e à FAPESP que mantem a pesquisa brasileira viva apesar de tudo. Que a sociedade saiba reconhecer valor dessas agências e a importância da Ciência para o desenvolvimento de uma nação.

“Todas as religiões, artes e ciências são ramos da mesma árvore. Todas essas aspirações se dirigem a enobrecer a vida do homem, levantando-a da esfera da mera existência física e conduzindo o indivíduo à liberdade”.

Albert Einstein (1879-1955)

RESUMO

PINTO, A. L. **Diagnóstico genético, gestão e visualização de dados genômicos de pacientes com síndromes de falência medular hereditária.** 2023. 109f. Tese (doutorado) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto; 2023.

As síndromes de falência medular hereditária (SFMHs) compõem um grupo heterogêneo de doenças ultrarraras caracterizadas por uma ou mais citopenias e hipoplasia da medula óssea. As SFMHs são consequências de variantes genéticas germinativas que resultam na diminuição prematura de populações de células-tronco hematopoéticas e no aumento da predisposição a neoplasias. O uso de tecnologias de sequenciamento genético em larga escala vem acelerando o diagnóstico dessas doenças e inundando os laboratórios de pesquisa com dados genômicos. Devido ao alto custo para geração da informação e a sua relevância científica, urge a necessidade de uma gestão de dados mais eficiente. O presente trabalho objetivou avaliar o perfil genético dos pacientes com suspeita de SFMHs e desenvolveu uma ferramenta para auxiliar no gerenciamento e na visualização de dados genômicos e laboratoriais, garantido segurança, integridade, acessibilidade e reutilização dessas informações. Primeiramente, foram sequenciados 314 pacientes com suspeitas de SFMHs por meio painel multigênico para realização do diagnóstico genético. Em seguida, foi modelado um banco de dados relacional para armazenar essas informações de forma segura e garantir a integridade dos dados. Para facilitar o acesso aos dados armazenados, desenvolvemos um aplicativo web (*BMF DataViz*) como ferramenta de visualização interativa das informações geradas pelo estudo. A análise genética identificou variantes em genes associados as SFMHs, em 46,4% dos pacientes (146). Foram encontradas variantes em genes da biologia dos telômeros (81), da biogênese ribossomal (27), de mecanismos de reparo (13) e de outras vias (25). Quase um terço (31,9%) das variantes priorizadas no estudo ainda não haviam sido descritas nos maiores bancos públicos de variantes genéticas. Esses resultados revelam como países com perfis genéticos pouco conhecidos podem contribuir para ampliar o conhecimento etiológico de doenças ultrarraras. O uso da ferramenta de armazenamento e visualização proposta permitirá o acesso a informações genômicas, laboratoriais e clínicas vindas de uma fonte única de dados. Com mais dados acessíveis será possível realizar análises mais robustas, como treinar algoritmos de inteligência artificial para auxiliar no manejo clínico desses pacientes.

Palavras-chaves: Síndromes de falência medular hereditária. Diagnóstico genético. Gestão e visualização de dados.

ABSTRACT

PINTO, A. L. **Genetic diagnosis, management and visualization of genomic data from patients with inherited bone marrow failure syndromes**. 2023. 109p. Thesis (PhD) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto; 2023.

Inherited bone marrow failure syndromes (IBMFS) are a heterogeneous group of ultra-rare diseases characterized by one or more cytopenias and bone marrow hypoplasia. IBMFS are the consequence of germline genetic variants that result in a premature decrease in hematopoietic stem cell populations and an increased predisposition to neoplasms. The use of large-scale genetic sequencing technologies has accelerated the diagnosis of these diseases and inundated research laboratories with data from rare genomes. Given the high cost of generating this information and its scientific relevance, there is an urgent need for more efficient data management. The aim of this study was to evaluate the genetic profile of patients with suspected IBMFS and to develop a tool to support the management of genomic and laboratory data, ensuring the security, integrity, accessibility and reusability of this information. Firstly, 314 patients with suspected IBMFS were sequenced using a multigenic panel for genetic diagnosis. Then, a relational database was modeled to store this information securely and ensure data integrity. To facilitate access to the stored data, we developed a web application (BMF DataViz) as a tool for visualizing the information generated by the study. The genetic analysis identified variants in genes associated with IBMFS in 46.4% of the patients (146). Variants were found in genes related to telomere biology (81), ribosomal biogenesis (27), repair mechanisms (13), and other pathways (25). Almost a third (31.9%) of the variants prioritized had not previously been described in the largest public genetic databases. The knowledge generated by the finding of these novel variants will contribute to the differential diagnosis of other patients with suspected IBMFS. These results reveal how countries with unknown genetic profiles can contribute to expanding the etiological knowledge of ultra-rare diseases. The use of storage and visualization tools proposed will allow access to genetic and laboratory information from a single source of truth (SSOT). With more accessible data, it will be possible to perform robust analyses, including training artificial intelligence algorithms to assist in the clinical management of these patients.

Keywords: Inherited bone marrow failure syndromes. Genetic diagnosis. Data management and visualization.

LISTA DE FIGURAS

Figura 1 - Volume de dados (em gigabytes) de sequenciamentos genéticos gerados pelo Laboratório Translacional de Hematologia.	20
Figura 2 – Exemplo do conceito de banco de dados relacional	22
Figura 3 – Exemplo de gráfico <i>lollipop</i> e o <i>oncoprint</i>	24
Figura 4 - Figura resumo de material e método	28
Figura 5 - Perfil das variantes genéticas filtradas.....	35
Figura 6 - Genes alterados em pacientes com diagnóstico genético de telomeropatias.....	36
Figura 7 - Genes alterados em pacientes com diagnóstico genético de ribossomopatia	37
Figura 8 - Genes alterados em pacientes com diagnóstico genético de anemia de Fanconi	37
Figura 9 - Genes alterados em pacientes com diagnóstico genético de outras SFMHs	38
Figura 10 - Diagrama da modelagem do banco de dados.	39
Figura 11 - Modelo conceitual do banco de dados centrado no paciente.	41
Figura 12 - Modelo lógico do banco de dados com relacionamentos.....	42
Figura 13 - Modelo físico do banco de dados.....	43
Figura 14 - Página inicial do aplicativo web <i>BMF DataViz</i>	44
Figura 15 - Página <i>Waterfall plot</i> do aplicativo web <i>BMF DataViz</i>	45
Figura 16 - Página <i>lollipop plot</i> do aplicativo web <i>BMF DataViz</i>	46
Figura 17 - Página <i>filtered variants</i> do aplicativo web <i>BMF DataViz</i>	46
Figura 18 - Página <i>telomere plot</i> do aplicativo web <i>BMF DataViz</i>	47
Figura 19 - Página <i>patient timeline</i> do aplicativo web <i>BMF DataViz</i>	48
Figura 20 - Diagrama de Venn com os genes alterados em diferentes estudos	52
Figura 21 - Gráfico de Pareto dos genes com alterações relevantes pelo número de pacientes.....	53

LISTA DE TABELAS

Tabela 1 – Painel gênico	30
--------------------------------	----

LISTA DE SIGLAS

AA - anemia aplástica

ACID - atonicidade, consistência, isolamento e durabilidade

ADB - anemia de Diamond-Blackfan

AF - anemia de Fanconi

BED - dado extensível do navegador (do inglês, *browser extensible data*)

CADD - do inglês, *combined annotation dependent depletion*

CH - cirrose hepática

CMD - conjunto mínimo de dados

COSMIC - *Catalogue of Somatic Mutations in Cancer*

CTH - célula-tronco hematopoética

DC - disceratose congênita

DEB - diepoxibutano

DR - doença rara

ERD - diagrama entidade relacionamento (do inglês, *Entity-Relationship Diagram*)

ETL – extração, transformação e carregamento (do inglês, *extract, transform and load*)

FAIR - localizável, acessível, interoperável e reutilizável (do inglês, *findable, accessible, interoperable, and reusable*)

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo

FP - fibrose pulmonar

FK – chave estrangeira (do inglês, *foreign key*)

HC - hematopoese clonal

HC-FMRP-USP - Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo

HC-FM-USP - Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo

HC-UFPR - Hospital das Clínicas da Universidade Federal do Paraná

IA - Inteligência Artificial

IDE - ambiente de desenvolvimento integrado (do inglês, *integrated development environment*)

Indel – variante genética do tipo inserção ou deleção

kb - quilobases

LGPD - Lei Geral de Proteção de Dados Pessoais

NCG - neutropenia congênita grave

ncRNA - RNA não-codificante (do inglês, *non-coding RNA*)

ng - nanograma

NGS - sequenciamento de nova geração (do inglês, *next generation sequencing*)

NHLBI-NIH - *Hematology Branch, National Heart, Lung, and Blood Institute at National Institutes of Health*

PEP - prontuário eletrônico do paciente

PGD - Planos de Gestão de Dados

PK - chave primária (do inglês, *primary key*)

PCR - Reação em cadeia da polimerase (do inglês, *polymerase chain reaction*)

RNA - ácido ribonucleico (do inglês, *ribonucleic acid*)

SDS - síndrome de Diamond-Shwachman

SFM - síndromes de falência de medular

SFMH - síndromes de falência de medular hereditária

SGBD - Sistema de gerenciamento de banco de dados

SHH - síndrome de Hoyeraal-Hreidarsson

SNV - variante de um único nucleotídeo (do inglês, *single nucleotide variant*)

SQL - linguagem de consulta estruturada (do inglês, *Structured Query Language*)

TAC - trombocitopenia amegacariocítica congênita

TCLE - termo de consentimento livre e esclarecido

TL - comprimento telomérico (do inglês, *telomere length*)

TS - sequenciamento direcionado (do inglês, *targeted sequencing*)

TMO - transplante de medula óssea

VCF - arquivo de chamada de variantes (do inglês, *variant calling file*)

VUS - variante de significado incerto (do inglês, *variant of uncertain significance*)

WGS - sequenciamento do genoma completo (do inglês, *whole-genome sequencing*)

WES - sequenciamento do exoma completo (do inglês, *whole-exome sequencing*)

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Ferramentas de auxílio ao diagnóstico de doenças raras e ultrarraras	15
1.2	Diagnóstico genético das síndromes de falência medular hereditária	16
1.3	Geração e gestão de dados para pesquisa de doenças genéticas raras e ultrarraras	19
1.4	Banco de dados	21
1.5	Visualização de dados	23
2	OBJETIVOS	26
2.1	Objetivo Geral	26
2.2	Objetivos Específicos	26
3	MATERIAL E MÉTODOS	28
3.1	Diagnóstico genético	28
3.1.1	Casuística	28
3.1.2	Coleta de dados e amostra biológica para extração de DNA	29
3.1.3	Preparo da biblioteca e sequenciamento massivo paralelo	29
3.1.4	Análise de variantes germinativa	31
3.2	Modelagem e instanciação do banco de dados	32
3.3	Aplicação web para acesso e visualização dos dados	32
4	RESULTADOS	35
4.1	Métricas do sequenciamento	35
4.2	Metade dos pacientes tinha alguma variante em genes associados as SFMHs	36
4.3	Modelagem de banco de dados	38
4.4	<i>BMF DataViz</i> : ferramenta interativa de visualização de dados genômicos, clínicos e laboratoriais	44
5	DISCUSSÃO	50
5.1	Diagnóstico genético de doenças raras e ultrarraras na população brasileira	50
5.2	Abordagens para diagnóstico genético das SFMHs	51
5.3	Como a gestão e compartilhamento de dados pode promover o avanço do conhecimento de doenças raras	54
5.4	Aplicação web interativa de visualização de dados genômicos, laboratoriais e clínicos de pacientes com SFMHs	55
6	CONCLUSÕES	57
7	SUGESTÕES PARA TRABALHOS FUTUROS	58

8 REFERÊNCIAS	60
APÊNDICE A - Genes associados a Síndrome de Falência de Medula Óssea hereditária	74
APÊNDICE B - Tabela de diagnóstico e alterações genéticas dos pacientes com suspeita de Síndrome de Falência de Medula Óssea hereditária	76
APÊNDICE C - Genes alterados em pacientes (N = 146) com diagnóstico genético de síndrome de falência medular hereditária.....	90
APÊNDICE D – Comprimento teloméricos de paciente com diagnóstico genético de telomeropatia	91
APÊNDICE E – Gráficos de <i>lollipops</i> dos dez genes com mais variantes	92
APÊNDICE F – Código em SQL do modelo físico	94
ANEXO 1 – Aprovação do Comitê de Ética em Pesquisa (CEP) em humanos	105

INTRODUÇÃO

1 INTRODUÇÃO

1.1 Ferramentas de auxílio ao diagnóstico de doenças raras e ultrarraras

Doenças raras (DRs) são enfermidades que afetam até 65 a cada 100 mil habitantes, segundo o Ministério da Saúde (BRASIL, 2014). No mundo, as DRs acometem de 263 a 466 milhões de pessoas (NGUENGANG WAKAP *et al.*, 2019). As doenças ultrarraras são ainda menos prevalentes, afetando 1 a cada 50 mil habitantes, segundo o *National Institute for Health and Care Excellence* (NICE) (NICE, 2004). A legislação brasileira define como doenças ultrarraras enfermidades crônicas, debilitantes ou que aumentam o risco de morte, com incidência menor que 1:50.000 (BRASIL, 2017).

A prevalência global das doenças ultrarraras vem aumentando exponencialmente em decorrência do surgimento de ferramentas modernas que auxiliam no diagnóstico (CROOKE, 2021). Ainda assim, estima-se que o tempo médio para diagnóstico dessas doenças é de 4,8 anos e que um terço dos pacientes morrem nos primeiros 5 anos de vida (LIU *et al.*, 2019). A falta de diagnóstico atrasa o tratamento correto e traz sofrimento psicológico para os pacientes e familiares.

Mais de 70% das doenças raras tem como fator etiológico alguma alteração genética (NGUENGANG WAKAP *et al.*, 2019). Apesar de ainda ser considerado de alto custo, o sequenciamento genômico vem se tornando cada vez mais acessível. Os testes genéticos podem diminuir a odisseia diagnóstica e reduzir gastos desnecessários para os serviços de saúde (LIU *et al.*, 2019).

Atualmente, o sequenciamento de nova geração (NGS) é amplamente utilizado no diagnóstico genético das DRs. Existem três estratégias que podem ser realizadas pelo método de NGS: o sequenciamento do genoma completo (WGS, do inglês, *whole-genome sequencing*), o sequenciamento do exoma completo (WES, do inglês, *whole-exome sequencing*) e o painel multigênico ou sequenciamento direcionado (TS, do inglês, *targeted sequencing*).

A relevância do diagnóstico genético para as DRs nem sempre está associada a opções terapêuticas. Após o resultado do teste, os médicos podem continuar tratando apenas os sintomas das doenças, porém com mais conhecimento sobre prognóstico daquele paciente (LIU *et al.*, 2019). Além disso, o diagnóstico genético

pode confortar a família, elucidando a causa da doença e auxiliando no planejamento familiar.

Os algoritmos de inteligência artificial (IA) já se apresentam como potenciais aliados em diversas áreas da saúde e podem ser usados no contexto das DRs (HAUG; DRAZEN, 2023; RAJKOMAR; DEAN; KOHANE, 2019; VISIBELLI *et al.*, 2023). No entanto, esses algoritmos necessitam de muitos dados para serem treinados e as doenças raras e ultrarraras têm escassez dessa matéria-prima (COHEN; MELLO, 2019; DECHERCHI *et al.*, 2021). Portanto, é essencial que pesquisadores de DRs cada vez mais compartilhem dados que auxiliem no desenvolvimento de algoritmos de IA com alta performance, trazendo benefícios clínicos para os pacientes.

1.2 Diagnóstico genético das síndromes de falência medular hereditária

As síndromes de falência medular (SFM), ou simplesmente falências medulares, compõem um grupo heterogêneo de doenças ultrarraras caracterizadas por uma ou mais citopenias em decorrência da insuficiência da medula óssea. A incidência de todas as SFMs ainda é desconhecida (DOKAL; TUMMALA; VULLIAMY, 2022).

As SFMs estão associadas a um risco aumentado de desenvolvimento de neoplasias e altas taxas de mortalidade (YOUNG; CALADO; SCHEINBERG, 2006). O protótipo das falências medulares é a anemia aplástica (AA), definida por pancitopenia com medula óssea hipocelular, na ausência de infiltrado ou fibrose medular (MARSH *et al.*, 2009). A AA é classificada como moderada, grave ou muito grave, de acordo com a celularidade da medula e as contagens do hemograma (BACIGALUPO *et al.*, 1988). As SFM podem ser hereditárias ou adquiridas (imune ou secundária).

Estima-se que até 30% das SFMs são causadas por alterações genéticas germinativas, grupo que inclui as AA constitucionais (SHIMAMURA; ALTER, 2010). As causas etiológicas das principais síndromes de falência medular hereditária (SFMHs) são variantes em genes relacionados a mecanismos de reparo do DNA, biologia dos telômeros e biogênese ribossomal, levando à diminuição da população de células-tronco hematopoéticas (CTHs) (DOKAL; VULLIAMY, 2010).

A SFMHs mais frequente é a anemia de Fanconi (AF), com incidência de 11,4 casos por milhão de nascimentos (TSANGARIS *et al.*, 2011). Etiologicamente, a AF é

um distúrbio de instabilidade cromossômica causado por variantes germinativas em, pelo menos, 21 genes de reparo de DNA da via FA/BRCA (Apêndice A). Atualmente, o diagnóstico laboratorial ainda é feito por meio do DEB teste, método baseado na exposição de células ao diepoxibutano (DEB), um agente indutor de quebra no DNA (ESMER *et al.*, 2004). Como as células dos pacientes com AF possuem defeitos nos mecanismos de reparo, observa-se um maior número de quebras cromossômicas em comparação ao controle, o que caracteriza um “DEB positivo”. Além disso, esses pacientes podem apresentar hipoplasia radial, baixa estatura, microcefalia e manchas café com leite (AUERBACH, 2009).

A disqueratose congênita (DC) é uma AA constitucional causada por variantes genéticas germinativas responsáveis pela manutenção e reparo dos telômeros. A DC clássica é diagnosticada pela presença da tríade mucocutânea (pigmentação reticulada da pele, displasia ungueal e leucoplasia oral) e os indivíduos afetados apresentam telômeros muito curtos para a idade (<1º percentil) (CALADO; YOUNG, 2009). Além da AA, pacientes com DC podem desenvolver fibrose pulmonar (FP), cirrose hepática (CH), estenose esofágica, uretral e dos ductos lacrimais (VULLIAMY; DOKAL, 2006). A forma mais grave da DC é conhecida como síndrome de Hoyeraal-Hreidarsson (SHH). Na SHH observa-se também retardo no crescimento intrauterino, microcefalia e hipoplasia cerebelar (DOKAL, 2000). A DC é o protótipo de um conjunto de doenças denominado de telomeropatias ou distúrbios da biologia dos telômeros (TOWNSLEY; DUMITRIU; YOUNG, 2014). Alterações germinativas em pelo menos 15 genes podem causar telomeropatias (Apêndice A). As telomeropatias têm penetrância e expressividade variável e, por isso, a maior parte dos pacientes desenvolvem apenas fenótipos de AA, FP ou CH.

Variantes patogênicas em pelo menos 20 genes relacionados à biogênese ribossomal estão associadas a duas SFMHs, a anemia de Diamond-Blackfan (ADB) e a síndrome de Diamond-Shwachman (SDS) (Apêndice A). Ambas são consideradas ribossomopatias, mas apresentam diferenças clínicas importantes. Pacientes com ADB desenvolvem anemia grave em consequência de uma hipoplasia na linhagem eritróide (VLACHOS *et al.*, 2008). Na SDS, os pacientes apresentam insuficiência pancreática exócrina e neutropenia que eleva os riscos de infecções (BOOCOOCK *et al.*, 2002).

Alterações genéticas germinativas em outros 8 genes já foram associadas a neutropenia congênita grave (NCG) (Apêndice A). Esses genes não estão relacionados a nenhum processo biológico em comum. Pacientes com NCG tipicamente apresentam infecções graves na infância, abscessos cutâneos e/ou pneumonia (WELTE; ZEIDLER, 2009).

A trombocitopenia amegacariocítica congênita (TAC) é um distúrbio ainda mais raro caracterizado pela ausência ou diminuição no número de precursores megacariocíticos na medula óssea e consequente plaquetopenia (BALLMAIER *et al.*, 2001). A TAC é causada por variantes patogênicas no gene *MPL*, que codifica o receptor da trombopoietina (Apêndice A).

Mais recentemente, foram descobertas outras SFMHs, como a deficiência de GATA2, MECOM, CTLA4, SAMD9 e LIG4 (Apêndice A). GATA2 é um importante fator de transcrição que atua na diferenciação e autorrenovação das CTHs. Variantes nesse gene causam a síndrome de deficiência de GATA2, clinicamente caracterizada por monocitopenia e infecção micobacteriana, fúngica ou viral (HSU *et al.*, 2013). Alteração no gene *MECOM* pode causar, além da falência medular, manifestações sistêmicas como sinostose radioulnar, clinodactilia, perda auditiva e malformações cardíacas e renais (GERMESHUSEN *et al.*, 2018). Variantes no gene *CTLA4* afetam a regulação imunológica, causando doenças intestinais, adenopatia, infecções e doenças autoimunes (KUEHN *et al.*, 2014). Mutações germinativas nos genes *SAMD9* e *SAMD9L* estão associadas à síndrome MIRAGE (um acrônimo de mielodisplasia, infecção, restrição de crescimento, hipoplasia adrenal, problemas genitais e enteropatia) (DAVIDSSON *et al.*, 2018). A síndrome de LIG4 (DNA ligase IV) é uma imunodeficiência primária, caracterizada por microcefalia, anomalias faciais e sensibilidade à radiação ionizante, causada por variantes no gene *LIG4* (ALTMANN; GENNERY, 2016).

Os pacientes com SFMHs têm em comum maior chance de evoluírem com neoplasias como síndrome mielodisplásica, leucemia mieloide aguda, leucemia linfóide aguda e tumores sólidos. Dependendo do gene, o padrão de herança pode ser autossômico dominante, recessivo, ou recessivo ligado ao X (Apêndice A). O transplante de medula óssea (TMO) ou terapias com hormônios sexuais são os tratamentos de primeira linha das falências medulares hereditárias. O surgimento dos sintomas nas SFMHs é variável. Pacientes com ADB ou NCG normalmente

apresentam desordens hematológicas no período neonatal, enquanto na AF e na DC, as citopenias podem se desenvolver mais tardiamente na infância ou em qualquer faixa etária (DOKAL; TUMMALA; VULLIAMY, 2022; WEGMAN-OSTROSKY; SAVAGE, 2017; YOUNG, 2018).

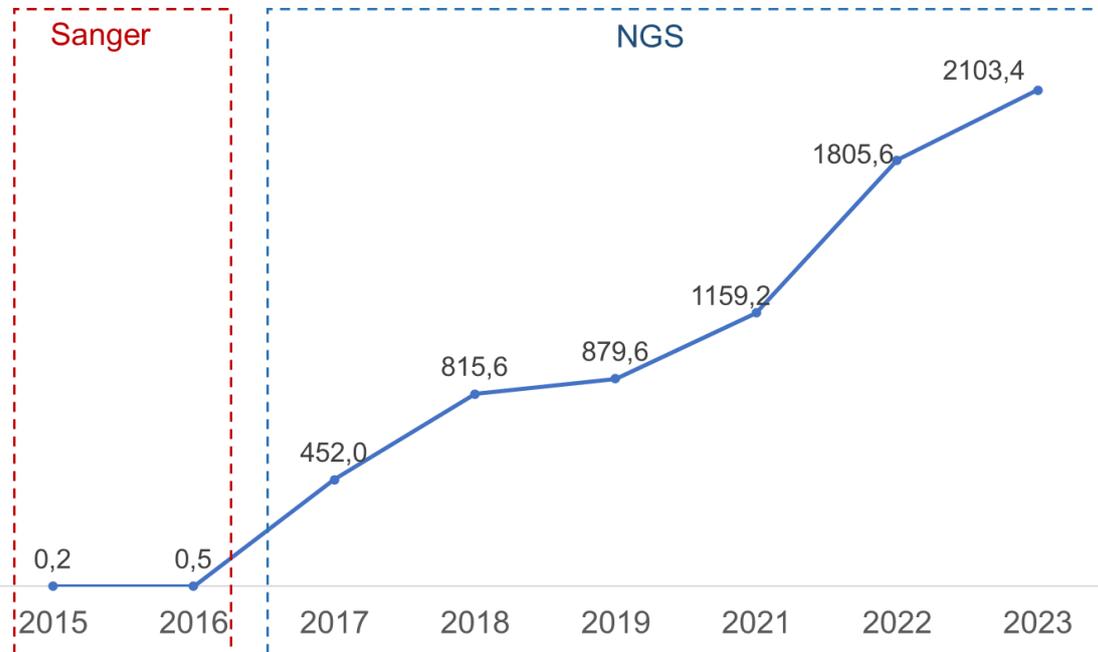
Atualmente, as estratégias de WES e painéis multigênicos são amplamente utilizados como ferramenta para o diagnóstico genético das SFMHs (BLUTEAU *et al.*, 2018; GÁLVEZ *et al.*, 2021; MURAMATSU *et al.*, 2017; TSANGARIS *et al.*, 2011). Os testes baseados em painéis oferecem alta sensibilidade devido à cobertura uniforme e profunda, sendo eficazes na detecção de variantes genéticas raras, lidando com a heterogeneidade genética e fenotípica da doença (LIU *et al.*, 2019). O WES, em geral, é usado quando os testes genéticos convencionais ou baseados em painel fornecem resultados inconclusivos. WES gera mais informações genéticas, possibilitando a identificação de variantes em genes até então não associados ao fenótipo (LIU *et al.*, 2019). No entanto, esse aumento na geração de dados acarreta mais complexidade para análise e pode não trazer benefícios clínicos em comparação com o painel multigênico (OUELLETTE *et al.*, 2018).

1.3 Geração e gestão de dados para pesquisa de doenças genéticas raras e ultrarraras

Assim como em outros setores, as pesquisas biomédicas vêm agregando tecnologias que aumentam a capacidade de mensurações dos fenômenos estudados. Em geral, essas novas ferramentas trazem um aumento na qualidade e, principalmente, na quantidade de informações geradas pelos pesquisadores (Figura 1). Com isso, surge a necessidade de uma gestão de dados mais eficiente por parte dos laboratórios e centros de pesquisa.

Figura 1 - Volume de dados (em gigabytes) de sequenciamentos genéticos gerados pelo Laboratório Translacional de Hematologia.

Aumento de **10.507 vezes** no volume de dados de sequenciamento (em gigabytes)



Sanger (sequenciamento pelo método de Sanger) e NGS (sequenciamento de nova geração)
 Fonte: Resultados originais da pesquisa.

Nos últimos anos, a academia, a indústria, as agências de fomento e as editoras de revistas científicas vêm discutindo e estabelecendo princípios para a gestão de dados científicos. O objetivo é tornar os dados científicos localizáveis, acessíveis, interoperáveis e reutilizáveis tanto para humanos como para máquinas. Esses princípios ficaram conhecidos pelo acrônimo FAIR, do inglês, *findable, accessible, interoperable, and reusable* (WILKINSON *et al.*, 2016).

Muitos pesquisadores de doenças raras trabalham com estudos de coorte prospectiva multicêntrica. Nesse tipo de estudo, são incluídos indivíduos de diversos centros, que podem estar em diferentes regiões e que devem ser acompanhados por anos. A raridade dessas informações reforça a necessidade da aplicação dos princípios FAIR no contexto das doenças raras (*Rare Diseases Global Open FAIR implementation Network*, 2019).

Durante esses trabalhos, é comum pesquisadores recorrerem a revisão manual de prontuários para extrair dados clínicos de pacientes (sinais, sintomas, diagnósticos, histórico de doenças, etc.) com auxílio de médico especialista, o que eleva o custo da coleta. Em seguida, geralmente os dados são armazenados em

tabelas criadas pelo próprio pesquisador (silos de dados). Isso resulta numa descentralização no armazenamento do dado, tornando-o pouco localizável. Além disso, a falta de padronização da informação nessas tabelas dificulta a reutilização do dado. Sendo assim, a cada novo projeto, os pesquisadores optam pela coleta de informações clínicas que, novamente, serão armazenados em mais uma tabela pouco localizável e com a reutilização comprometida.

Ademais, a criação desses silos de dados impede o controle de acesso e compromete a segurança de informações sensíveis dos pacientes. A Lei Geral de Proteção de Dados Pessoais (LGPD) diz que é responsabilidade dos centros de pesquisas manter os dados “em ambiente controlado e seguro, conforme práticas de segurança previstas em regulamento específico e que incluam, sempre que possível, a anonimização ou pseudonimização dos dados, bem como considerem os devidos padrões éticos relacionados a estudos e pesquisas” (BRASIL, 2018).

Para a coleta e armazenamento de dados clínicos, ferramentas como REDCap (HARRIS *et al.*, 2009), quando bem aplicadas, atendem as demandas dos princípios FAIR e da LGPD. No entanto, armazenar as saídas das análises genômicas permitindo buscas mais eficientes e precisas, pode ser um desafio. Cada exoma, por exemplo, gera uma tabela com 40-50 mil linhas e centenas de colunas. Ao customizar um banco relacional para atender as demandas de armazenamento de dados genômicos estruturados, é possível realizar buscas e análises de forma mais performática, além de permitir escalabilidade.

1.4 Banco de dados

Dados são unidades individuais de informação. Em outras palavras, dados são “fatos conhecidos que podem ser registrados e que possuem significado implícito” (ELMASRI; NAVATHE, 2019). Existem dados do tipo estruturado (e.g. dados tabulares); semiestruturado, (e.g. formato XML, do inglês, *extensible markup language*); e não-estruturado (e.g. imagens, áudios e textos). Os dados consumidos por uma aplicação, como um prontuário eletrônico, precisam estar armazenados em um banco de dados de forma organizada.

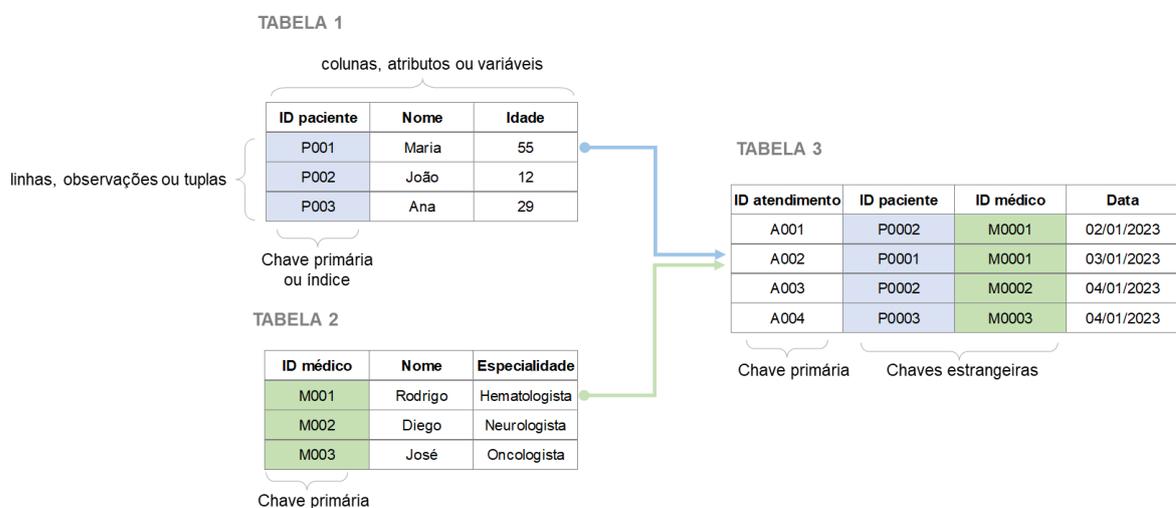
Um banco de dados é um conjunto de dados relacionados, caracterizado: (1) por ser uma representação de uma porção da realidade, também chamada de minimundo; (2) por ser uma coleção lógica de dados com algum significado inerente;

(3) e por ser projetado, construído e preenchido com valores para um propósito específico, com usuários e aplicações predefinidas (ELMASRI; NAVATHE, 2019).

O processo de construção de um banco de dados passa pela criação dos modelos conceitual, lógico e físico. O modelo conceitual é a primeira etapa do processo e nesta é feita a descrição do contexto de negócio, num alto nível de abstração. Em seguida, desenha-se o modelo lógico descrevendo as estruturas que estarão contidas no banco de dados, como os relacionamentos. Por último, o modelo físico descreve a estrutura física de armazenamento de dados, definindo os tipos de dados, os tamanhos dos campos, os índices, entre outros aspectos. (MACHADO, 2020).

Existem bancos de dados relacionais e não-relacionais, que armazenam dados estruturados e semiestruturados, respectivamente. Os bancos de dados relacionais se baseiam num modelo criado por Codd na década de 70 e são muito utilizados até hoje (CODD, 1970). Um banco de dados relacional é composto por várias tabelas interoperáveis entre si (Figura 2). As linhas das tabelas são tuplas (sequência de valores) referentes às observações. As colunas são os atributos que dão sentido aos valores das linhas. O estabelecimento de relações entre as tabelas se dá por meio de um identificador único, também conhecido como índice ou chave-primária (PK, do inglês, *primary key*). Uma linha de uma tabela pode se relacionar com uma ou mais linhas de outra tabela.

Figura 2 – Exemplo do conceito de banco de dados relacional



Fonte: Resultados originais da pesquisa.

A gestão de um banco relacional é feita por um sistema de gerenciamento de banco de dados (SGBD), que garante as propriedades transacionais, conhecidas pelo acrônimo ACID: atomicidade, consistência, isolamento e durabilidade (GRAY, 1981). Os SGBDs dos bancos de dados relacionais usam linguagem de consulta estruturada (SQL, do inglês *structured query language*). Os comandos DDL (do inglês, *data definition language*) no SQL são usados, por exemplo, para criar, alterar e apagar tabelas. Já os comandos DML (do inglês, *data query language*) realizam consultas, inserem, deletam e atualizam dados. Além disso, outras funções do SQL são usadas para controlar acesso, garantir integridade e compartilhar dados (ELMASRI; NAVATHE, 2019).

Atualmente, SGBDs como o *MySQL Workbench* (Oracle) contam com uma interface gráfica bastante amigável. No entanto, acessar dados por comandos SQL ainda é uma barreira para os usuários comuns. Portanto, aplicações de visualização de dados são indispensáveis para permitir o acesso a grande quantidade de informações armazenadas em um banco relacional.

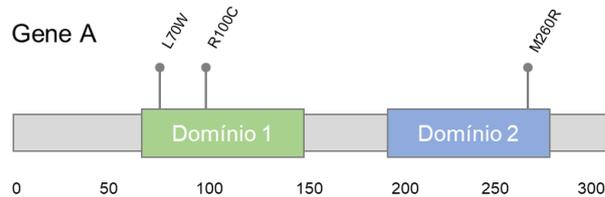
1.5 Visualização de dados

A visualização de dados (ou DataViz, do inglês, *data visualization*) emergiu como campo de pesquisa nos anos 2000, resultado da fusão de duas subáreas da Ciência de Computação: a visualização científica (ou SciVis, do inglês, *scientific visualization*) e a visualização de informação (ou InfoVis, do inglês, *information visualization*) (O'DONOGHUE *et al.*, 2018).

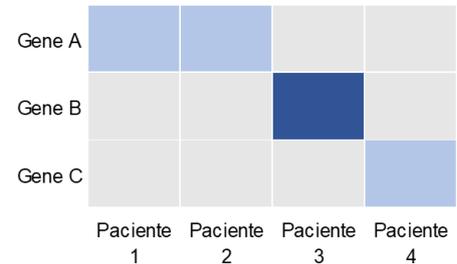
As ferramentas de visualização desempenham papéis fundamentais na exploração de dados gerados pelas “ômicas” (genômica, transcriptômica, epigenômica, entre outras) (WU *et al.*, 2022). Como exemplo, temos os gráficos *lollipop* e o *oncoprint*, duas estratégias de visualização bastante utilizadas para facilitar a interpretação e comunicação de variantes genéticas (Figura 3). Ambas foram popularizadas pelo portal de visualização e análise de dados genômicos de câncer, o cBioPortal (CERAMI *et al.*, 2012; GAO *et al.*, 2013).

Figura 3 – Exemplo de gráfico *lollipop* e o *oncoprint*

A – *Lollipop plot*



B – *Oncoprint*



Fonte: Resultados originais da pesquisa.

Outro desafio de visualização ocorre quando os pesquisadores são obrigados a condensar grande quantidade de dados genômicos em imagens de dimensões reduzidas para a publicação. Além de pequenas, as figuras de artigos científicos são estáticas, o que impossibilita os leitores de explorarem detalhes de seu interesse (PERKEL, 2018). Os gráficos interativos são uma alternativa para melhorar o acesso à grande quantidade de dados genômicos gerados pelos estudos.

A visualização interativa compõe um conjunto de ferramentas que possibilitam que o leitor filtre e amplie gráficos, altere a forma como os dados são representados e acesse informações subjacentes (OLA; SEDIG, 2018). Os pacotes *Plotly* e *Shiny* são, respectivamente, exemplos de ferramentas de geram e compartilham visualizações interativas (CHANG *et al.*, 2023; INC., 2015). Periódicos do grupo *Nature* e *Cell* vêm dando cada vez mais espaço as figuras interativas (DUNHAM *et al.*, 2012; PERKEL, 2018; RAO *et al.*, 2017).

Dessa forma, o presente trabalho visa dar acesso rápido e interativo a grande quantidade de dados genômicos, laboratoriais e clínicos dos pacientes com SFMHs, gerados em anos de pesquisa. Isso facilitará a comunicação dos dados, permitirá reanálises mais frequentes e auxiliará na geração de hipóteses para novos projetos.

OBJETIVOS

2 OBJETIVOS

2.1 Objetivo Geral

Conhecer o perfil genético de pacientes brasileiros com suspeita de síndromes de falência medular hereditária e desenvolver uma ferramenta que auxilie no gerenciamento e permita visualização de dados genômicos e laboratoriais, garantido segurança, integridade, acessibilidade e reutilização das informações geradas por esse estudo.

2.2 Objetivos Específicos

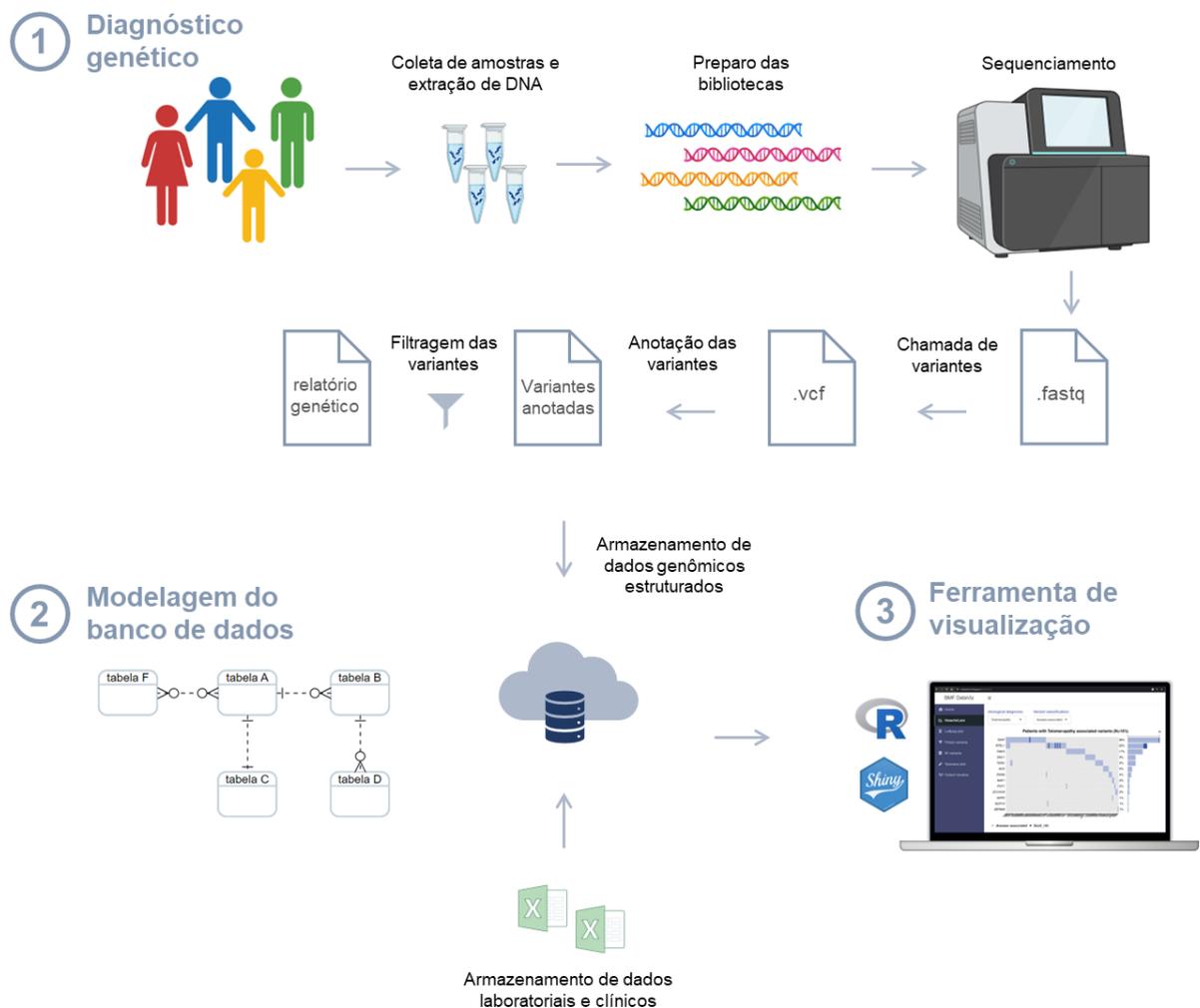
- Realizar diagnóstico genético de pacientes com suspeita de síndromes de falência medular hereditária;
- Modelar banco de dados para gerenciamento de dados genômicos, clínico e laboratoriais dos pacientes;
- Implementar ferramenta interativa de visualização dos dados gerados pelo estudo.

MATERIAL E MÉTODOS

3 MATERIAL E MÉTODOS

A Figura 4 resume as etapas do diagnóstico genético, da modelagem do banco de dados e da ferramenta de visualização desenvolvida.

Figura 4 - Figura resumo de material e método



Fonte: Algumas imagens da figura foram criadas pelo biorender.com

3.1 Diagnóstico genético

3.1.1 Casuística

Foram incluídos no estudo pacientes (N = 314) do sexo feminino (N = 147) e masculino (N = 167), não relacionados, com suspeita de telomeropatias (telômero curto para idade, uma ou mais citopenias e medula hipocelular ou aplásica), de ribossomopatias (aplasia pura de série vermelha) e de outras síndromes de falências de medular hereditária (uma ou mais citopenias e medula hipocelular ou aplásica) com

ou sem histórico familiar. Pacientes que preenchiam os critérios de inclusão, mas que apresentaram resposta à imunossupressão, DEB teste positivo e/ou clone HPN positivo foram excluídos do estudo.

Os pacientes do estudo são acompanhados por especialistas em falência medular no Hospital das Clínicas da Faculdade de Medicina da USP de Ribeirão Preto (HC-FMRP-USP), no Hospital das Clínicas da Faculdade de Medicina da USP de São Paulo (HC-FM-USP), no Hospital das Clínicas da Universidade Federal do Paraná (HC-UFPR), no Hospital Pequeno Príncipe de Curitiba e na Beneficência Portuguesa de São Paulo.

A concordância em participar da pesquisa se deu por meio da assinatura do Termo de Consentimento Livre e Esclarecido (TCLE) aprovado pelo Comitê de Ética em Pesquisa (CEP) do HC-FMRP-USP (nº 4.940.397) (Anexo 1). Após o consentimento do paciente, ou do seu responsável legal, foram coletadas amostras biológicas (sangue periférico) e dados clínicos e laboratoriais.

3.1.2 Coleta de dados e amostra biológica para extração de DNA

Os dados clínicos e laboratoriais dos pacientes foram coletados através dos Sistemas de Informação Hospitalar do HC-FMRP-USP ou através do formulário de envio de amostras de participantes de outros centros.

Foram coletadas amostras de sangue periférico (4 mL) dos pacientes e os leucócitos totais foram isolados para extração de DNA. O DNA total foi extraído com a utilização do kit comercial *DNeasy Blood & Tissue Kit (Qiagen)*, seguindo as instruções do fabricante. A quantificação foi feita no Qubit 3.0 (*Thermo Fisher Scientific*), as razões 260/280 e 260/230 foram observadas no nano-espectrofotômetro *NanoDrop2000 (Thermo Fisher Scientific)* e a integridade do DNA foi verificada em gel de agarose (0,8%). Para o sequenciamento foram utilizadas as amostras de DNA que apresentaram razões 260/280 e 260/230 entre 1,8 e 2,1, e banda única no gel.

3.1.3 Preparo da biblioteca e sequenciamento massivo paralelo

As bibliotecas foram preparadas utilizando o kit de captura *SureSelect QXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library (Agilent Technologies)* customizado com painel de 169 genes (Tabela1).

Para o preparo das bibliotecas, foram utilizados 50 ng de amostra de DNA genômico de cada paciente. O DNA das amostras foi fragmentado enzimaticamente

em sequências de 150 pares de bases. Foram adicionados adaptadores às extremidades dos fragmentos seguido por purificação com *beads* magnéticas *AMPure XP Beads* (Agencourt).

Os fragmentos com os adaptadores foram hibridados a sondas de RNA complementares aos genes de interesse (biblioteca de enriquecimento por captura híbrida). As moléculas hibridadas foram capturadas e minimamente amplificadas (10 ciclos). Por fim, uma sequência específica (*barcode* ou índice) foi adicionada aos fragmentos amplificados, servindo como identificador único para a separação *in silico* das amostras. As bibliotecas foram quantificadas por fluorimetria no Qubit 3.0 (*Thermo Fisher Scientific*), qualificadas por eletroforese capilar automatizada no *Fragment Analyzer* (*Agilent*) e normalizadas para sequenciamento massivo paralelo.

O sequenciamento foi realizado na plataforma *NextSeq 500* (*Illumina*) pelos métodos de “*Bridge PCR*” e “*SBS*” (*Sequencing-by-Synthesis*), utilizando o *NextSeq 500/550 High Output Kit v2* para 300 ciclos (*Illumina*).

Tabela 1 – Painel multigênico desenhado para síndromes de falência medular hereditária

<i>ABCA3</i>	<i>CDAN1</i>	<i>FANCC</i>	<i>JAK2</i>	<i>PHF6</i>	<i>RUVBL1</i>	<i>TERF2IP</i>
<i>ABL1</i>	<i>CDKN2A</i>	<i>FANCD2</i>	<i>JAK3</i>	<i>PIGA</i>	<i>RUVBL2</i>	<i>TERT**</i>
<i>ACD</i>	<i>CEBPA</i>	<i>FANCE</i>	<i>KDM6A</i>	<i>PML</i>	<i>SAMD9</i>	<i>TET2</i>
<i>ALAS2</i>	<i>CSF3R</i>	<i>FANCF</i>	<i>KIT</i>	<i>POT1</i>	<i>SAMD9L</i>	<i>THPO</i>
<i>ASXL1</i>	<i>CTC1</i>	<i>FANCG</i>	<i>KMT2A</i>	<i>PRF1</i>	<i>SBDS</i>	<i>TINF2</i>
<i>ATM</i>	<i>CTLA4</i>	<i>FANCI</i>	<i>KRAS</i>	<i>PTEN</i>	<i>SBF2</i>	<i>TP53</i>
<i>ATRX</i>	<i>CUX1</i>	<i>FANCL</i>	<i>LIG4</i>	<i>PTPN11</i>	<i>SETBP1</i>	<i>U2AF1</i>
<i>BCOR</i>	<i>CXCR4</i>	<i>FANCM</i>	<i>MECOM</i>	<i>RAD21</i>	<i>SF3B1</i>	<i>U2AF2</i>
<i>BCORL1</i>	<i>DAXX</i>	<i>FAS</i>	<i>MPL</i>	<i>RAD51</i>	<i>SFTPA2</i>	<i>UBE2T</i>
<i>BLM</i>	<i>DCLRE1B</i>	<i>FASLG</i>	<i>MRE11A</i>	<i>RAD51C</i>	<i>SFTPBB</i>	<i>USB1</i>
<i>BRAF</i>	<i>DDX11</i>	<i>FBXW7</i>	<i>MYC</i>	<i>RBM8A</i>	<i>SFTPC</i>	<i>VPS45</i>
<i>BRCA1</i>	<i>DDX41</i>	<i>FLT3</i>	<i>MYD88</i>	<i>RECQL4</i>	<i>SH2B3</i>	<i>WAS</i>
<i>BRCA2</i>	<i>DHX36</i>	<i>G6PC3</i>	<i>NAF1</i>	<i>RPL11</i>	<i>SHQ1</i>	<i>WRAP53</i>
<i>BRIP1</i>	<i>DKC1</i>	<i>GATA1</i>	<i>NBN</i>	<i>RPL15</i>	<i>SLX4**</i>	<i>WRN</i>
<i>CALR</i>	<i>DNAJC21</i>	<i>GATA2*</i>	<i>NHP2</i>	<i>RPL26</i>	<i>SMC1A</i>	<i>WT1</i>
<i>CBL</i>	<i>DNMT3A</i>	<i>GFI1</i>	<i>NOP10</i>	<i>RPL35A</i>	<i>SMC3</i>	<i>XRCC2</i>
<i>CBLB</i>	<i>EFL1</i>	<i>GNAS</i>	<i>NOTCH1</i>	<i>RPL5</i>	<i>SRP72</i>	<i>ZBTB48</i>
<i>CBLC</i>	<i>ELANE</i>	<i>GRHL2</i>	<i>NPM1</i>	<i>RPS10</i>	<i>SRSF2</i>	<i>ZCCHC8</i>
<i>CCT2</i>	<i>ERCC4</i>	<i>HAX1</i>	<i>NRAS</i>	<i>RPS17</i>	<i>STAG2</i>	<i>ZRSR2</i>
<i>CCT3</i>	<i>ERCC6L2</i>	<i>HRAS</i>	<i>OBFC1</i>	<i>RPS19</i>	<i>STAT3</i>	SNP rs35705950 (<i>MUC5B</i>)
<i>CCT4</i>	<i>ETV6</i>	<i>IDH1</i>	<i>PALB2</i>	<i>RPS24</i>	<i>TCP1</i>	SNP rs2736100 (<i>TERT</i>)
<i>CCT5</i>	<i>EZH1</i>	<i>IDH2</i>	<i>PARN</i>	<i>RPS26</i>	<i>TEN1</i>	SNP rs111521887 (<i>TOLLIP</i>)
<i>CCT6A</i>	<i>EZH2</i>	<i>IKZF1</i>	<i>PCNA</i>	<i>RPS7</i>	<i>TERC</i>	SNP rs5743890 (<i>TOLLIP</i>)
<i>CCT7</i>	<i>FANCA</i>	<i>IRF1</i>	<i>PDGFRA</i>	<i>RTEL1</i>	<i>TERF1</i>	SNP rs5743894 (<i>TOLLIP</i>)

CCT8 FANCB JAK1 PEG3 RUNX1 TERF2

Nota: *com intron 5, **com promotor

3.1.4 Análise de variantes germinativa

Foram utilizados para análise os arquivos *fastq*, gerados no sequenciamento, e o arquivo BED referente às regiões de interesse. A aplicação *DRAGEN Enrichment* v.3.10.4 da plataforma *BaseSpace (illumina)* foi utilizada na configuração padrão para o alinhamento e chamada de variantes. Utilizou-se o genoma de referência GRCh37/hg19 (*UCSC hg19 alt-masked v2*).

Os arquivos VCF (do inglês, *Variant Call Format*) gerados foram anotados utilizando o wAnnotar (1000G: 2015aug; ExAC: exac03; esp: esp6500siv2; dbSNP: avsn147; cosmic: cosmic81; clinvar: 20170501; gwas: avgwas_20170516; dbNSFP: dbnsfp33a; gnomad: gnomad_exome, gnomad_genome) (YANG; WANG, 2015).

As variantes germinativas foram filtradas e priorizadas seguindo os critérios:

- Estar em regiões exônicas, de *splicing* ou de RNA não-codificante;
- Ser não-sinônimas;
- Apresentar frequência populacional máxima menor ou igual a 0,01%;
- Ter score CADD (*combined annotation dependent depletion*) maior ou igual a 15;
- Ser lida com profundidade maior ou igual a 15;
- Apresentar qualidade da base maior ou igual a 20;
- Ter frequência alélica (AF) maior ou igual a 0,2;
- Aparecer em menos de dois registros no COSMIC;
- Estar genes associados a achados laboratoriais e/ou clínico.

As variantes filtradas foram classificadas como patogênica (P), provavelmente patogênica (LP, do inglês *likely pathogenic*), de significado incerto (VUS, do inglês, *variant of uncertain significance*), provavelmente benigna (LB, do inglês, *likely benign*) e benigna (B), seguindo os critérios da ACMG (RICHARDS *et al.*, 2015) e através das plataformas VarSome (KOPANOS *et al.*, 2019) e Franklin (<https://franklin.genoox.com/>).

3.2 Modelagem e instanciação do banco de dados

Primeiramente, mapeou-se o processo de aquisição de amostras, coleta e geração de dados do Laboratório Translacional de Hematologia (HC-FMRP-USP). O diagrama do mapeamento foi elaborado utilizando a aplicação Bizagi Modeler (versão 3.9.0.015). Em seguida, usando um alto nível de abstração, foi estabelecido um modelo conceitual do banco de dados centrado no paciente. O diagrama desse modelo também foi elaborado no programa Bizagi Modeler (versão 3.9.0.015). O modelo lógico do banco, que apresenta um nível intermediário de abstração, foi desenhado com a ferramenta online *VisualParadigm*.

Por fim, elaborou-se o modelo físico do banco de dados usando o MySQL Workbench (versão 8.0.34). Nele, foram estabelecidos os nomes das tabelas e colunas, os tipos de dados de cada coluna (inteiro, decimal, texto e data), as restrições (quais campos permitem dados nulos), as chaves primárias (PK, do inglês, *primary key*), as chaves estrangeiras (FK, do inglês, *foreign key*) e as cardinalidades (relacionamento entre as linhas das tabelas). Para garantir o estabelecimento de algumas relações foram criadas automaticamente tabelas-pivô.

O banco de dados MySQL foi hospedado num servidor em nuvem *Amazon Web Services* (AWS) alocado em São Paulo. Para isso foi utilizado o serviço de banco de dados relacional da Amazon (Amazon RDS, do inglês, *Amazon Relational Database Service*). Após instanciação do banco algumas tabelas foram povoadas com dados de arquivos csv, usando o *Table Data Import Wizard* no MySQL Workbench.

3.3 Aplicação web para acesso e visualização dos dados

A aplicação web para a visualização de dados genéticos e laboratoriais foi desenvolvida na linguagem de programação R, utilizando a versão 4.0.3 (R CORE TEAM, 2020) e o IDE RStudio/Posit (versão 2023.06.1) (RSTUDIO TEAM, 2020).

Para conexão com o banco de dado MySQL, foi usado o pacote RMySQL (versão 0.10.23). No caso de alguns dados laboratoriais que estavam no RedCAP, fizemos conexão com o banco via API, utilizando a biblioteca redcapAPI(versão 2.8.0).

Para o desenvolvimento do *back-end* e o *front-end* da aplicação, foram utilizados os pacotes Shiny (versão 1.7.2) (CHANG *et al.*, 2023), ShinyDashboardPlus (versão 2.0.3) (GRANJON, 2021) e ShinyDashboard (versão 0.7.2) (CHANG; BORGES RIBEIRO, 2021).

Para manipular tabelas e tratamento de dados, foram usadas as bibliotecas tidyR (versão 1.2.0) (WICKHAM; VAUGHAN; GIRLICH, 2023), dplyR (versão 1.1.3) (WICKHAM *et al.*, 2023), Janitor (versão 2.1.0) (FIRKE, 2023), StringR (versão 1.4.0) (GAGOLEWSKI, 2022) e Stringi (versão 1.7.8) (GAGOLEWSKI, 2022).

Para elaboração de gráficos e tabelas, foram utilizados Ggplot2 (versão 3.3.6) (WICKHAM, 2016), Plotly (versão 4.10.0) (INC., 2015), quantreg (versão 5.94) (KOENKER, 2022), MAtools (versão 2.6.05) (MAYAKONDA *et al.*, 2018), TimeVis (versão 2.1.0) (ATTALI; ALMENDE B.V., 2022) e DT (versão 0.23) (XIE; CHENG; TAN, 2022).

A aplicação Shiny foi hospedada no *Shiny Server*. Para controle de acesso foi adicionado uma autenticação, utilizando o pacote auth0 (versão 0.2.3) (FALBEL, 2023).

RESULTADOS

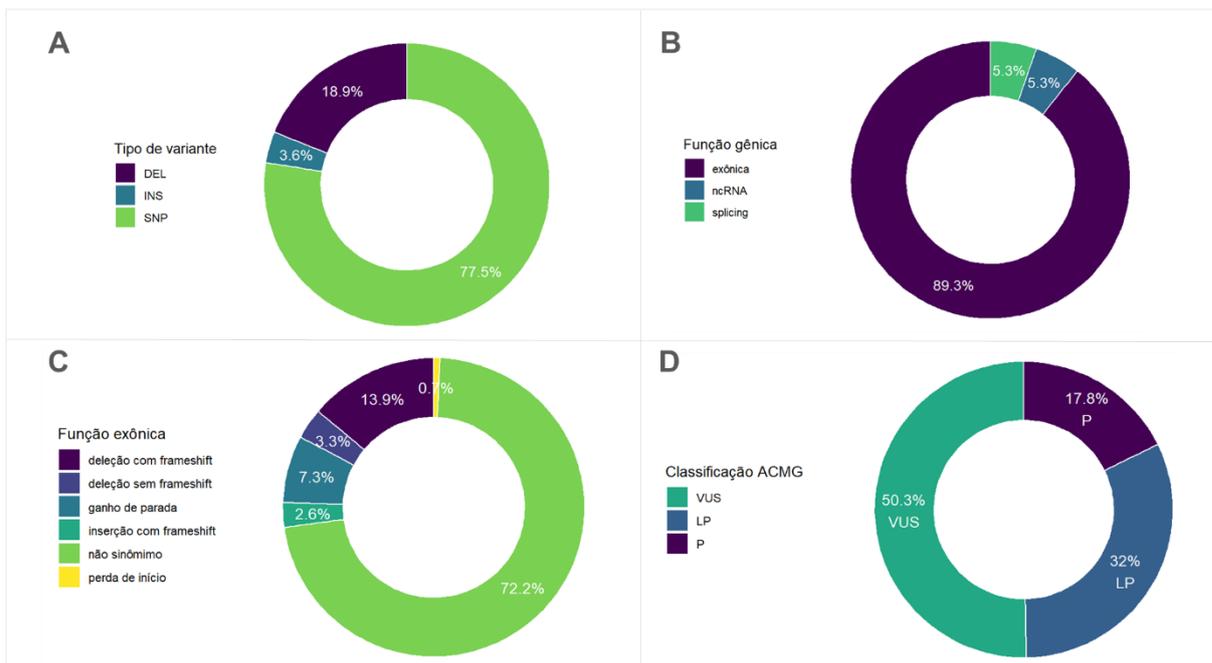
4 RESULTADOS

4.1 Métricas do sequenciamento

Mais de 370 mil variantes genéticas foram identificadas com o painel multigênico. Em média, 945 variantes de um único nucleotídeo (SNVs) e 165,5 inserções ou deleções (indels) foram identificadas por amostra. A cobertura vertical média (profundidade) foi de 349,7 vezes, e 99,3% da região-alvo foi coberta pelo menos 15 vezes.

Após a aplicação dos filtros, foram priorizadas 169 variantes de 146 pacientes (Apêndice B e C). Mais de dois terços das variantes foram SNPs (77,5%) e 22,5% foram *indels* (Figura 5 - A). A maioria das variantes foi exônica (89,3%), sendo 72,2% não-sinônimas (Figura 5 - B e C). Quanto à patogenicidade, 49,7% foram classificadas como patogênicas (P) ou provavelmente patogênicas (LP) (Figura 5 - D).

Figura 5 - Perfil das variantes genéticas filtradas



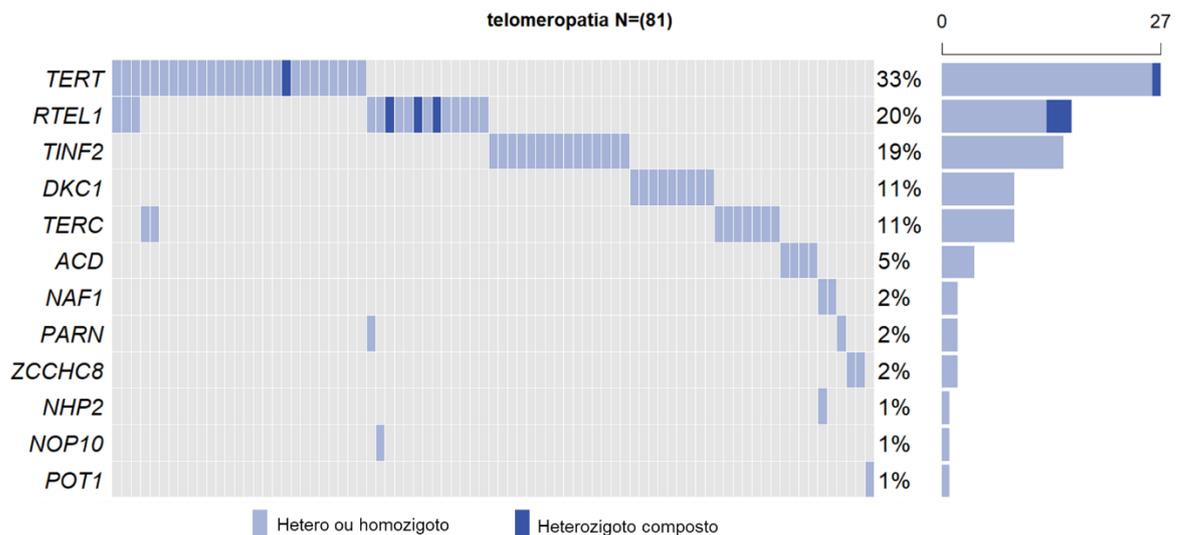
Fonte: Resultados originais da pesquisa.

Abreviação: Inserção (INS), deleção (DEL), variante de um único nucleotídeo (SNV, do inglês, *single nucleotide variant*), RNA não-codificante (ncRNA, em inglês, *non-coding RNA*), variante patogênica (P), provavelmente patogênica (LP, do inglês *likely pathogenic*) e VUS (do inglês, *variant of uncertain significance*)

4.2 Metade dos pacientes tinha alguma variante em genes associados as SFMHs

Em 46,4% (146) dos pacientes sequenciados foram encontradas variantes associadas às síndromes de falência medular hereditária. Dos 146 pacientes, 55% (81) tinham uma ou mais variantes (P, LP e VUS) nos genes associados a telomeropatias (Figura 6): *TERT* (27), *RTEL1* (16), *TINF2* (16), *DKC1* (9), *TERC* (9), *ACD*(4), *NAF1* (2), *PARN* (2) , *ZCCHC8* (2), *NHP2* (1), *NOP10* (1) e *POT1* (1) (Apêndice B, D e E).

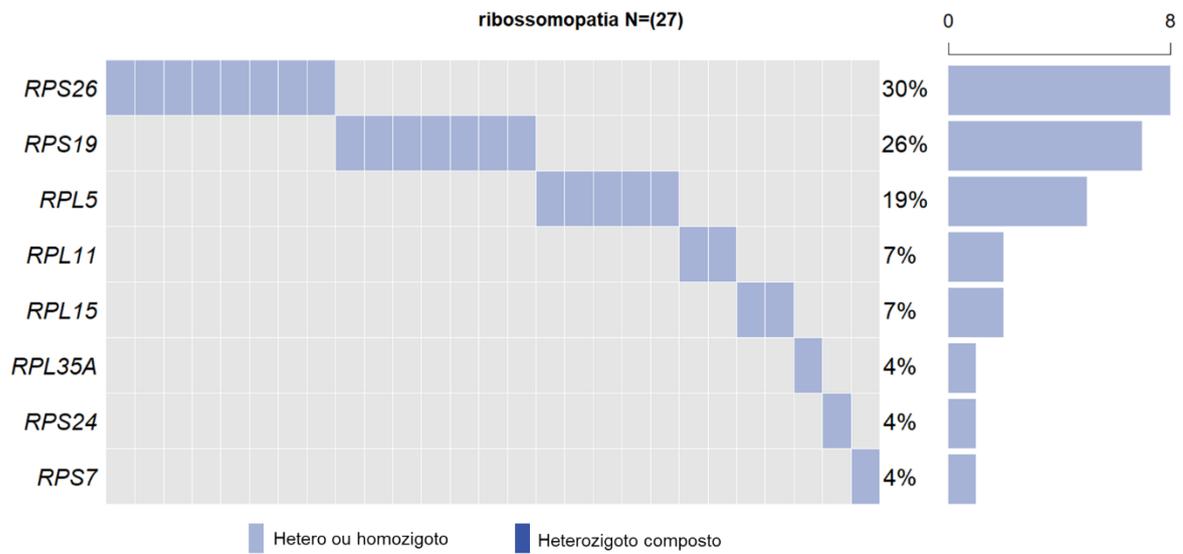
Figura 6 - Genes alterados em pacientes com diagnóstico genético de telomeropatias



Fonte: Resultados originais da pesquisa.

Na Figura 7 estão representados os 27 (18%) pacientes com alterações em genes ligados a ribossomopatia: *RPS26* (8), *RPS19* (7), *RPL5* (5), *RPL11* (2), *RPL15* (2), *RPL35A* (1), *RPS24* (1) e *RPS7* (1) (Apêndice B e E).

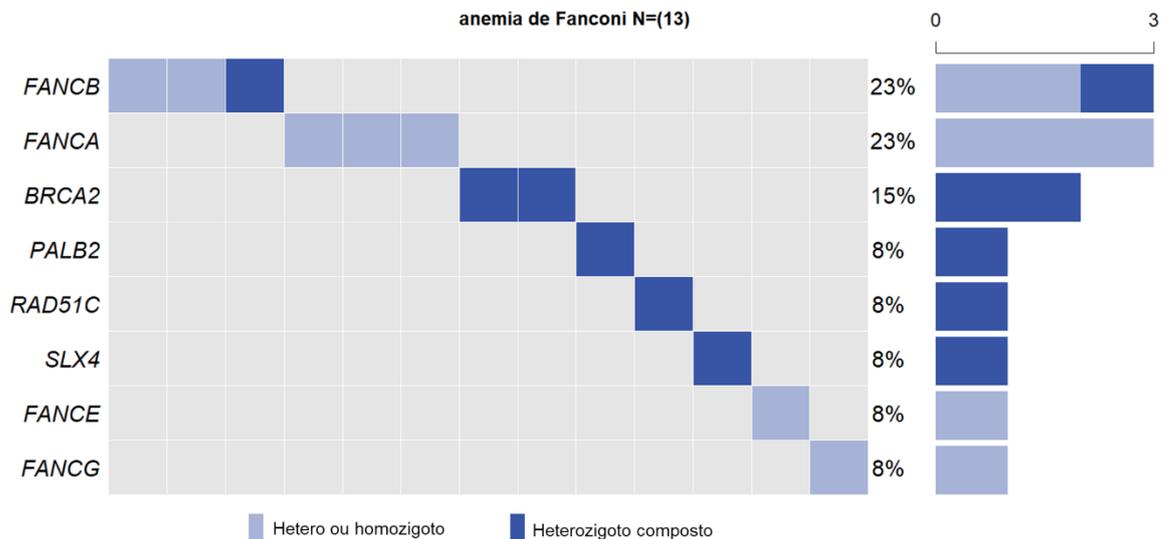
Figura 7 - Genes alterados em pacientes com diagnóstico genético de ribossomopatia



Fonte: Resultados originais da pesquisa.

Nove por cento (13) apresentaram alterações bialélicas patogênicas nos genes de reparo associados a anemia de Fanconi (Figura 8): *FANCA* (3), *FANCB* (3), *BRCA2* (2), *FANCE* (1), *FANCG* (1), *PALB2* (1), *RAD51C* (1) e *SLX4* (1) (Apêndice B e E).

Figura 8 - Genes alterados em pacientes com diagnóstico genético de anemia de Fanconi

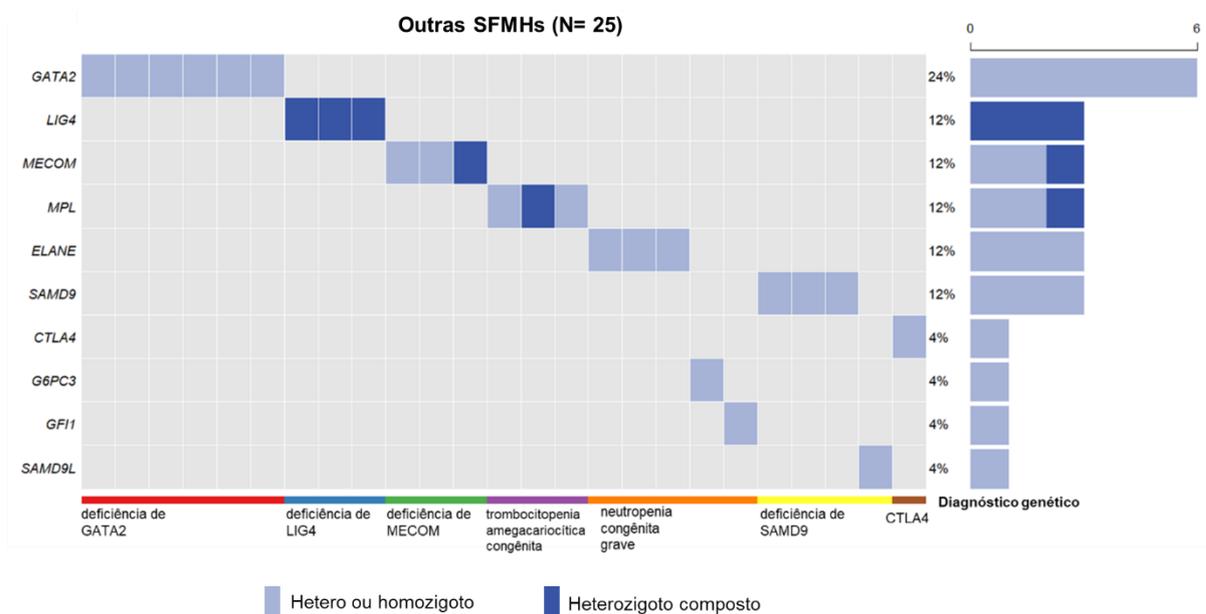


Fonte: Resultados originais da pesquisa.

Em 25 pacientes (17%) foram encontradas variantes associadas a outras SFMHs como deficiência de *GATA2* (6), *LIG4* (3), *MECOM* (3), *SAMD9* (5), *CTLA4* (1), trombocitopenia amegacariocítica congênita (3) e neutropenia congênita grave (5) (Figura 9). Encontramos variantes associadas a essas SFMHs nos genes *GATA2* (6),

LIG4 (3), *MECOM* (3), *ELANE* (3), *G6PC3* (1) e *GFI1* (1), *SAMD9* (3), *SAMD9L* (1) e *CTLA4* (1) (Apêndice B e E).

Figura 9 - Genes alterados em pacientes com diagnóstico genético de outras SFMHs



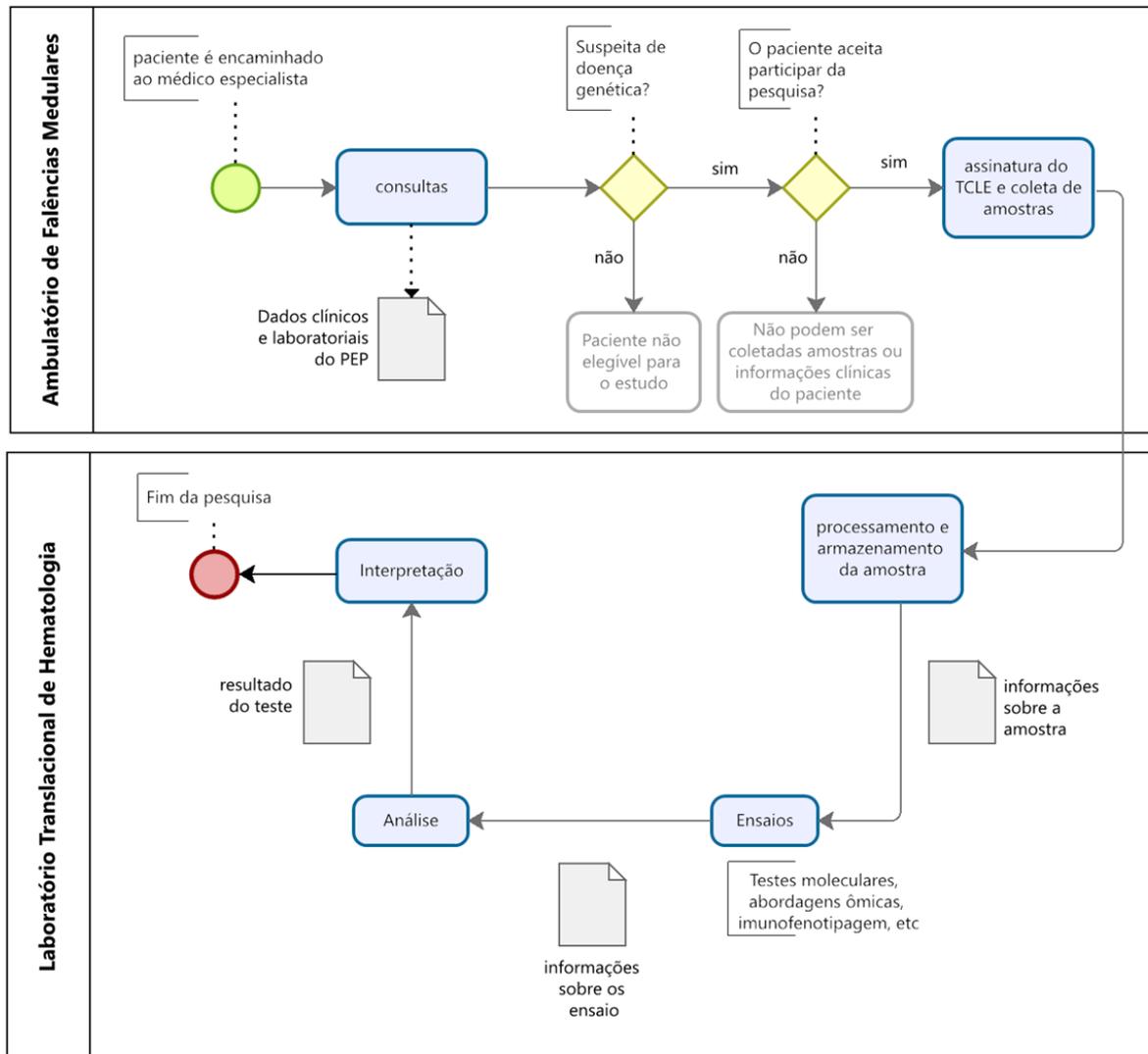
Outras síndromes de falência medular hereditária (SFMHs): deficiência de GATA2, MECOM, CTLA4, LIG4, SAMD9 e trombocitopenia amegacariocítica congênita e neutropenia congênita grave. Fonte: Resultados originais da pesquisa.

Quase um terço (32%) de todas as alterações encontradas no estudo ainda não haviam sido descritas nos maiores bancos públicos de variantes genéticas. Não foram encontradas variantes associadas as síndromes de falência medular hereditária em 53% (168) dos pacientes sequenciados (Apêndice B).

4.3 Modelagem de banco de dados

O diagrama do processo (Figura 10) mostra os quatro pontos de coleta/registo de informações e seus respectivos responsáveis: (1) consultas ambulatoriais (médico especialista), (2) processamento das amostras (técnico de laboratório ou pesquisador), (3) análise do resultado do ensaio (analista de laboratório ou pesquisador) e (4) interpretação do resultado do ensaio (médico e pesquisador).

Figura 10 - Diagrama da modelagem do banco de dados.



As etapas 1, 2, 3 e 4 do processo são os principais pontos de coleta/registo de informações.

Nota: PEP (prontuário eletrônico do paciente)

Fonte: Resultados originais da pesquisa.

Consideramos o início do processo como sendo, o momento em que o paciente é encaminhado a um centro de referência onde será acompanhado por médicos especialistas na enfermidade rara.

A primeira etapa de coleta de dados é realizada pelo médico especialista. Nas primeiras consultas, são coletadas a maior parte das informações clínicas, sociodemográficas e histórico familiar do paciente. Esses dados geralmente são registrados em textos narrativos no prontuário, que pode ser físico ou eletrônico.

Se confirmada a suspeita de doença genética, o paciente é convidado a entrar no estudo. Caso aceite, assinará um TCLE. Após esta autorização, o pesquisador pode coletar informações do prontuário, que comumente é o documento mais rico e

mais difícil de se extrair dados de maneira automatizada. Em geral, os pesquisadores revisam manualmente esses registros e transferem para uma planilha as informações que interessam para o seu projeto.

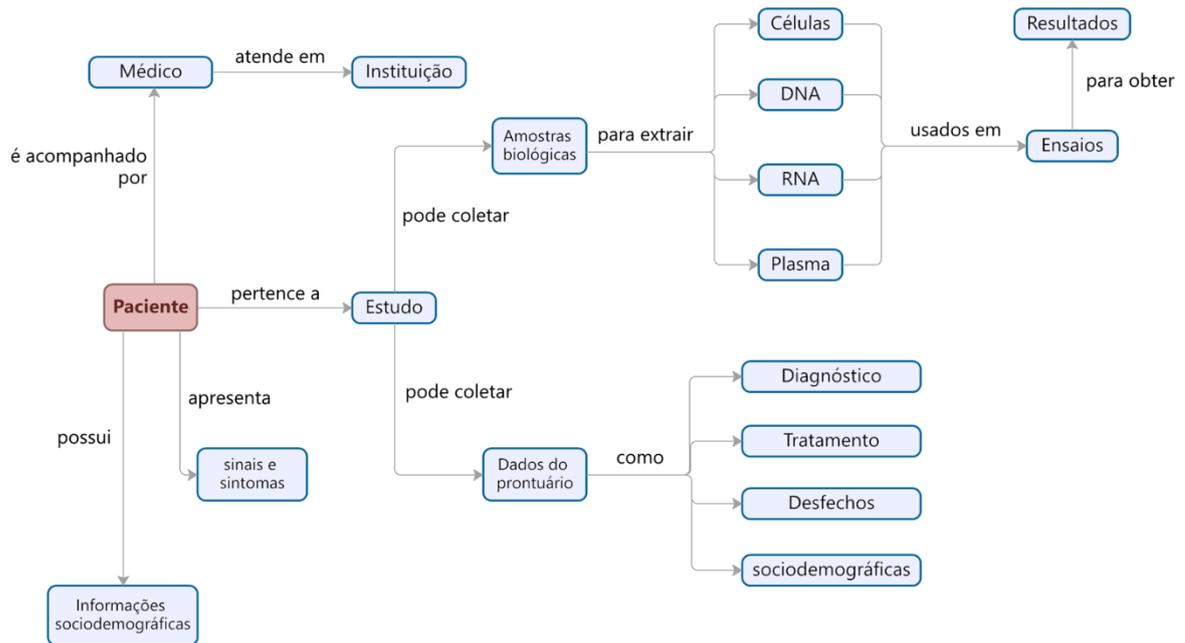
Após a assinatura do TCLE, além dos dados clínicos, podem ser coletadas amostras biológicas, como sangue, *swab* bucal ou pele do paciente. Essas amostras são levadas ao laboratório, onde são processadas e os subprodutos (células, DNA, RNA, proteínas, soro etc.) são armazenados. Nesse momento, são registradas informações como data da coleta, tipo de amostra, processamento e local de armazenamento. Este é o segundo ponto de coleta de dados.

A terceira etapa de coleta de dados, ocorre após as amostras serem submetidas a ensaios laboratoriais para confirmação da hipótese diagnóstica. Em seguida, os resultados obtidos são qualificados, antes de seguir para interpretação de uma equipe multiprofissional. Com o aumento na utilização de testes genéticos modernos, essa etapa do processo pode ser bastante impactada pelo grande volume de dados gerados nos ensaios. O ensaio de NGS, por exemplo, gera tabelas com dezenas de milhares de linhas e centenas de colunas por amostra.

A quarta e última etapa de registro de informações, em geral, é realizada por uma equipe multidisciplinar composta por médicos especialistas, pesquisadores e/ou analistas de laboratório. Nessa etapa, é feita a interpretação dos resultados obtidos nos ensaios, cruzando com os achados clínicos do paciente.

Após o mapeamento do processo e identificação dos pontos de coleta de dados, foi possível elaborar o modelo conceitual ou modelo entidade e relacionamento (Figura 11). O modelo conceitual tem um alto nível de abstração (MACHADO, 2020), pois o objetivo é apenas estabelecer as relações entre as entidades. O paciente é o ponto central desse modelo, pois a gestão moderna de dados na saúde é centrada no indivíduo e não nos processos.

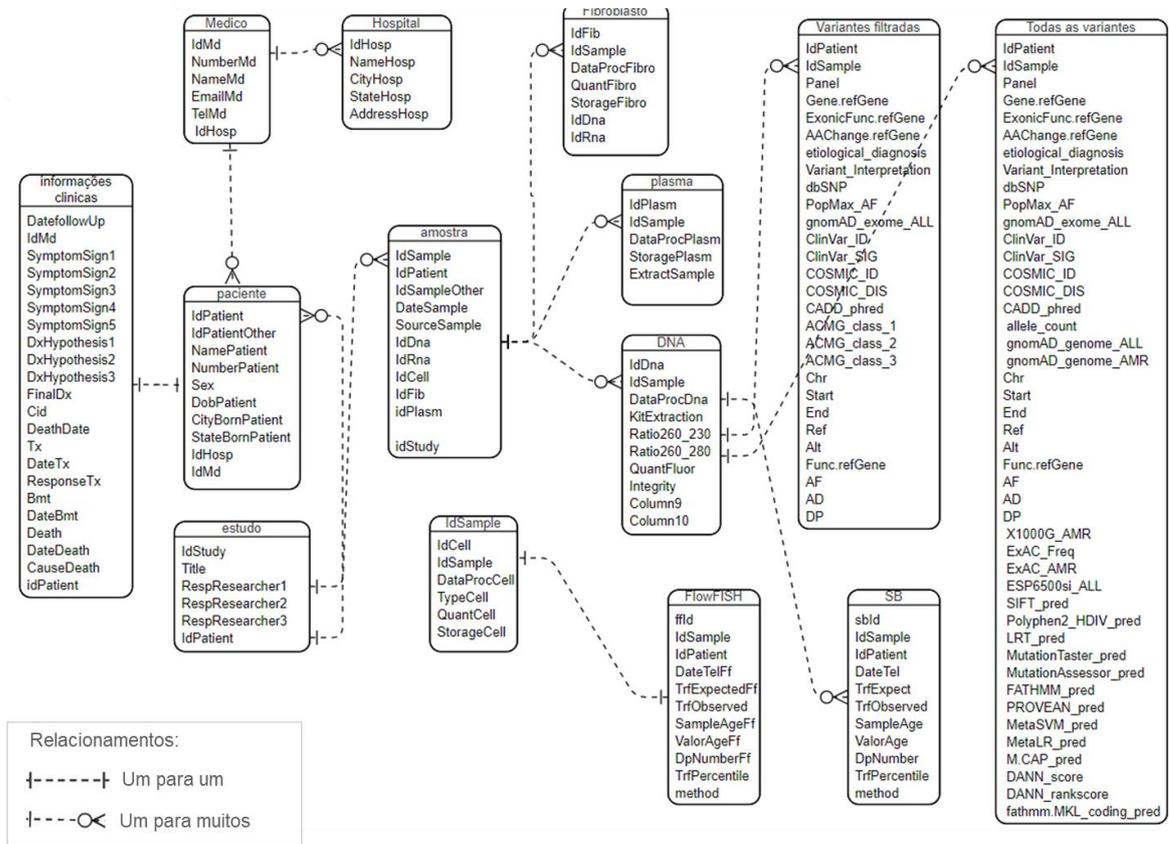
Figura 11 - Modelo conceitual do banco de dados centrado no paciente.



Fonte: Resultados originais da pesquisa.

Com estabelecimento das entidades e suas relações, foi possível definir a estrutura das tabelas, a cardinalidade e os atributos (Figura 12). No total, foram definidos 212 atributos (colunas) distribuídos em 14 tabelas para armazenar informações sociodemográficas, assistenciais, clínicas, laboratoriais e genéticas dos pacientes.

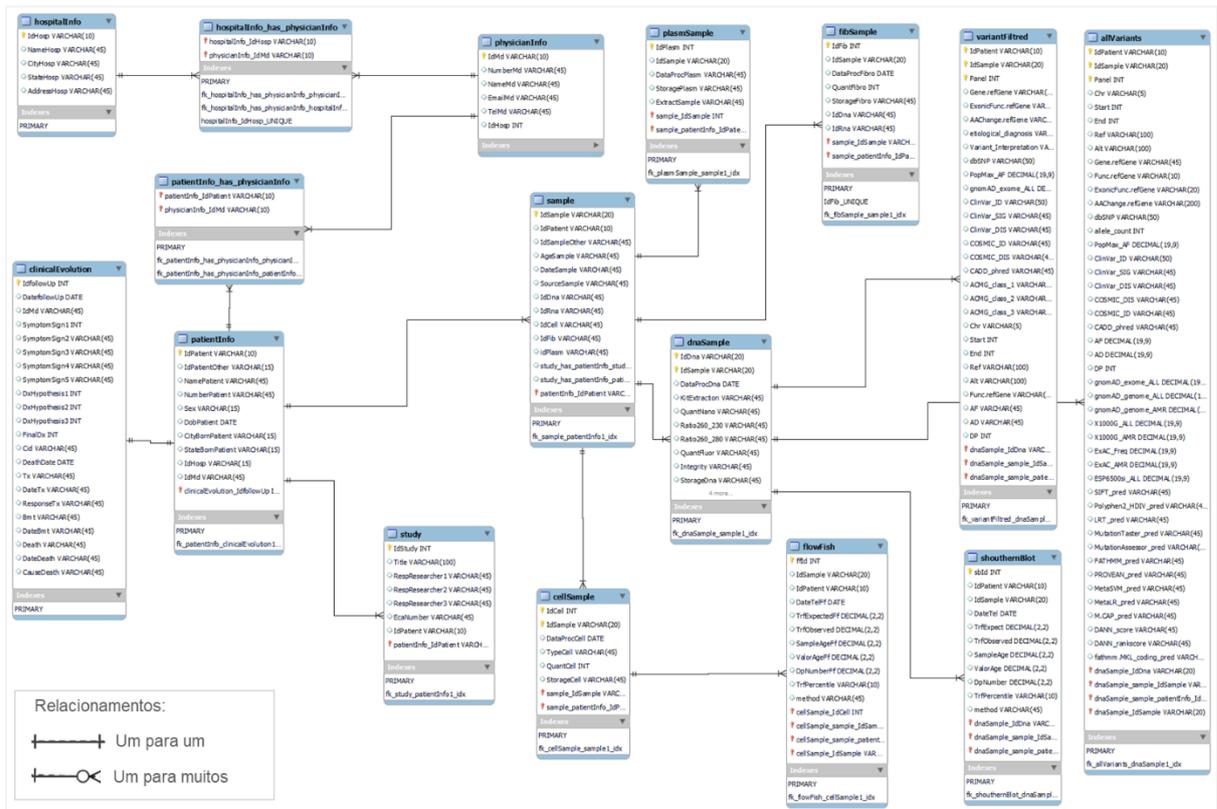
Figura 12 - Modelo lógico do banco de dados com relacionamentos.



Fonte: Resultados originais da pesquisa.

Definidos os nomes das tabelas e as colunas, partimos para o modelo físico (Figura 13). Esse modelo se baseou num formato de diagrama amplamente usado, o ERD (do inglês, *entity-relationship diagram*). As tabelas e colunas foram nomeadas em inglês para facilitar o compartilhamento com colaboradores estrangeiros. O código em SQL para elaboração do modelo se encontra no Apêndice F.

Figura 13 - Modelo físico do banco de dados



Fonte: Resultados originais da pesquisa.

O modelo físico do banco de dados apresenta uma arquitetura genérica, permitindo que o banco seja adaptado para usos mais específicos com a adição de poucas tabelas. Caso o pesquisador queira armazenar parâmetros hematológicos, por exemplo, uma tabela com dados de hemogramas completos (número de hemácias, leucócitos, plaquetas, etc) pode ser adicionada. No caso de laboratórios que trabalham com doenças metabólicas, pode ser relevante adicionar uma tabela contendo dados de enzimas hepáticas (fosfatase, transaminases, albumina, etc.). Para as pesquisas com doenças respiratórias, pode ser necessária uma tabela para os dados de espirometria (capacidade pulmonar total, capacidade vital, volume residual etc).

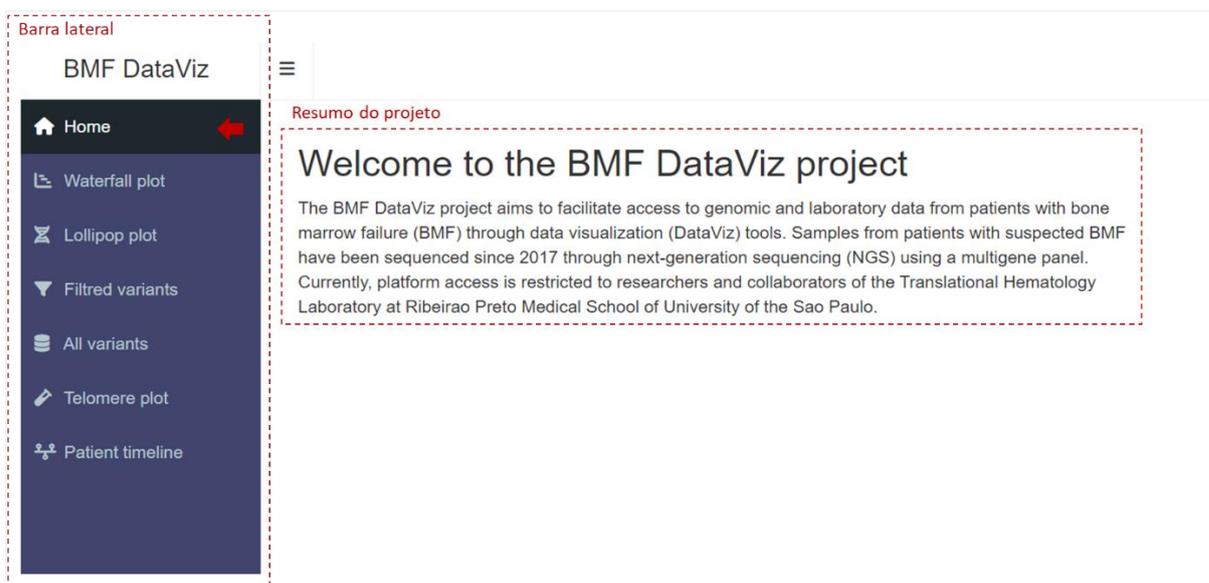
Os ganhos com o estabelecimento de um SGBD são inúmeros. O MySQL, por exemplo, cria um ambiente seguro, garante controle de acesso, evita redundância, facilita compartilhamento, realiza *backups* e permite restaurações do banco de dados de pesquisas. Além disso, o SGBD padroniza a entrada dos dados, o que facilita os processos de ETL (do inglês, *extract, transform and load*), garantindo mais reprodutibilidade da pesquisa e permitindo futuras integrações com outros bancos.

4.4 **BMF DataViz: ferramenta interativa de visualização de dados genômicos, clínicos e laboratoriais**

Desenvolvemos um aplicativo web interativo que recebeu o nome de *BMF DataViz* (do inglês, *bone marrow failure data visualization*), pois se utiliza de ferramentas de visualização de dados para facilitar a compreensão de grande quantidade de informação genética, clínica e laboratorial dos pacientes com falência medular. Com os elementos gráficos utilizados, pode-se resumir grandes quantidades de informação do banco de dados, facilitando a gestão, comunicação e interpretação.

O aplicativo conta com sete páginas, abas ou painéis. Na página inicial (*home*) é apresentada uma descrição do projeto em inglês (Figura 14). Na barra lateral esquerda, encontra-se um menu retrátil onde há opções para acessar as demais páginas. Toda ação que é realizada na interface do usuário (*front-end*) é executada no servidor da aplicação (*back-end*) trazendo novas informações.

Figura 14 - Página inicial do aplicativo web *BMF DataViz*.

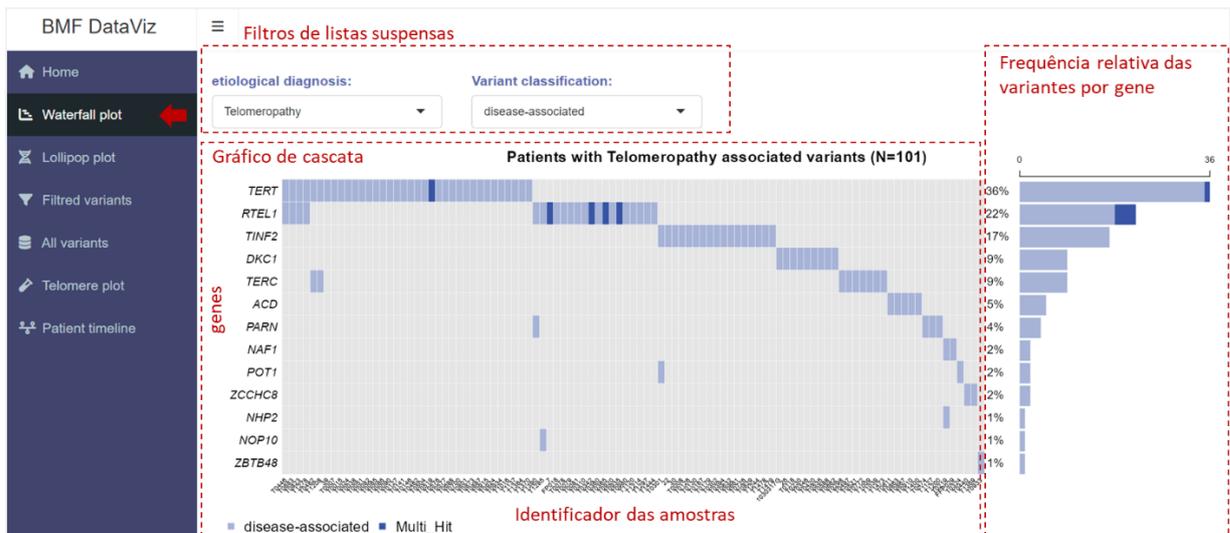


Fonte: Resultados originais da pesquisa.

A página *waterfall plot* apresenta gráficos do tipo *oncoprint*, comumente utilizados nos trabalhos de análise de variantes em amostras tumorais (Figura 15). O *oncoprint* é um mapa de calor que exhibe alterações genéticas em uma matriz de genes por amostras/pacientes. Nessa matriz os genes estão nas linhas e os pacientes em colunas. As cores podem ser usadas para representar diferentes tipos de alterações, mas nesse caso se refere a presença de múltiplas variantes em um mesmo gene de

um paciente. Na lateral direita, encontra-se um gráfico de barra com frequências relativas das alterações por genes. Nessa página, há filtros como listas suspensas que permitem que o usuário altere o gráfico, filtrando por diagnóstico etiológico (*Telomeropathy; Ribosomopathy; Fanconi anaemia; SAMD9; GATA2, MECOM, LIG4 and CTLA4 Deficiencies and CAMT; ICUS*) e classificação das variantes (*disease-associated; rare variant; somatic*).

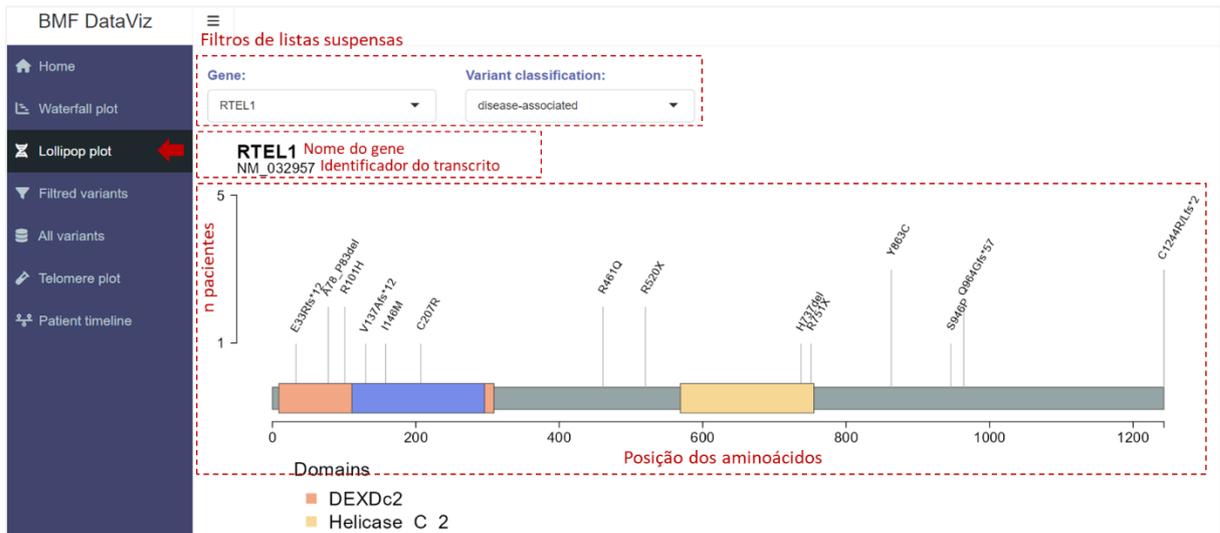
Figura 15 - Página *Waterfall plot* do aplicativo web *BMF DataViz*.



Fonte: Resultados originais da pesquisa.

A página *lollipop plot* apresenta gráficos de “pirulito” a partir da escolha de um gene e da classificação da variante por listas suspensas (Figura 16). O objetivo do gráfico *lollipop* é permitir que variantes sejam visualizadas juntamente com os domínios proteicos. Dessa forma é possível correlacionar o domínio alterado com diferentes fenótipos de doenças. De outra forma, representação gráfica mostra a relação entre uma variável categórica no eixo x (posição do aminoácido na proteína) e uma variável numérica y (quantidade de pacientes com a alteração).

Figura 16 - Página *lollipop plot* do aplicativo web *BMF DataViz*



Fonte: Resultados originais da pesquisa.

Na página *filtered variants* temos a lista de centenas de variantes que passaram pelos critérios de priorização (Figura 17). Esta página tem opções de filtrar as variantes por paciente, por interpretação da variante, por gene alterado, por identificador da variante em banco público (dbSNP), por diagnóstico etiológico e por ensaio. Após utilizar a estratégia de filtragem, o usuário pode baixar um arquivo .xlsx clicando em *download* e analisar em um leitor de planilha digital, como Microsoft Excel.

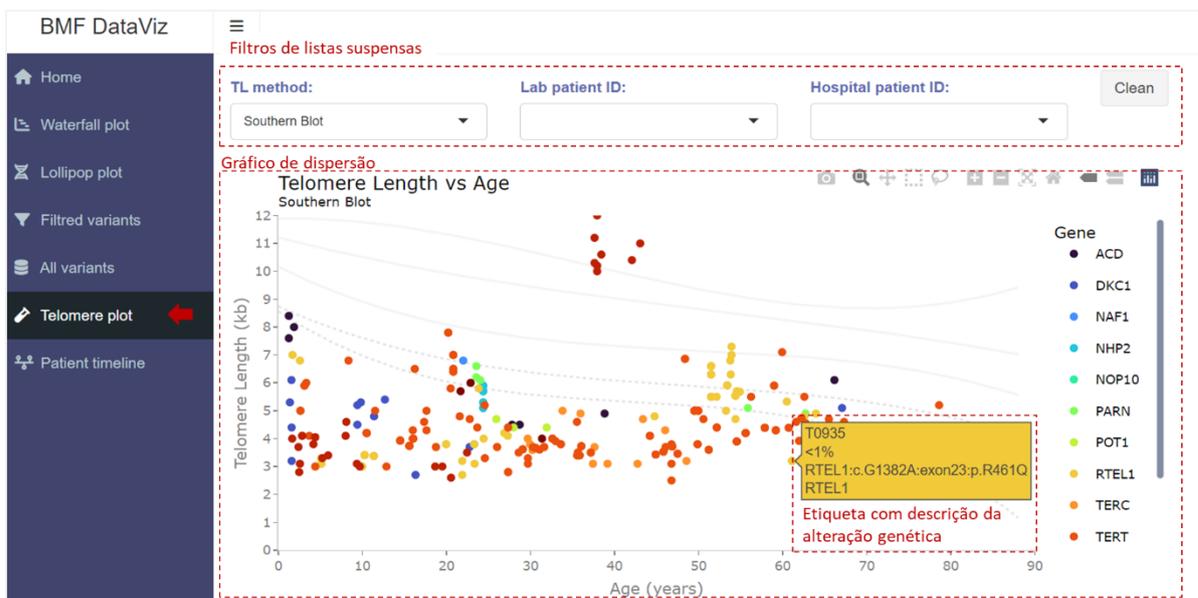
Figura 17 - Página *filtered variants* do aplicativo web *BMF DataViz*

PatientID	Gene.refGene	ExonicFunc.refGene	AAChange.refGene	GT	Variant Interpretation	etiologica
T0234	U2AF2	nonsynonymous SNV	U2AF2:NM_001012478:exon2:c.C87G:p.H29Q U2AF2:NM_007279:exon2:c.C87G:p.H29Q	0/1	rare variant	ICUS
T0273	SF3B1	nonsynonymous SNV	SF3B1:NM_012433:exon6:c.C608T:p.T203I	0/1	rare variant	SAMD9L
T0273	SAMD9	nonsynonymous SNV	SAMD9:NM_001193307:exon2:c.T866A:p.F289Y SAMD9:NM_017654:exon3:c.T866A:p.F289Y	0/1	disease-associated	SAMD9L
T0287	NOTCH1	nonsynonymous SNV	SF3B1:NM_012433:exon6:c.C608T:p.T203I	0/1	rare variant	ICUS
T0295	FBXW7	nonsynonymous SNV	FBXW7:NM_001013415:exon2:c.A212G:p.K71R FBXW7:NM_018315:exon2:c.A326G:p.K109R FBXW7:NM_033632:exon3:c.A566G:p.K189R FBXW7:NM_001349798:exon5:c.A566G:p.K189R	0/1	rare variant	Fanconi an
T0613					none	ICUS
T0645	SLX4	nonsynonymous SNV	SLX4:NM_032444:exon12:c.G4010T:p.R1337M	0/1	rare variant	Fanconi an

Fonte: Resultados originais da pesquisa.

Na página *telomere plot* encontra-se um gráfico de dispersão (comprimento telomérico por idade e gene alterado) interativo. Ao passar o cursor sobre qualquer ponto, surgirá uma etiqueta com a descrição da alteração genética (Figura 18). Na lista suspensa *TL method* é possível escolher o método de mensuração do telômero (*Southern blot* ou *Flow-FISH*), com isso as curvas serão alteradas. Os pontos, correspondendo a cada paciente, são visualizados sobre curvas derivadas da distribuição de comprimento telomérico por idade em indivíduos saudáveis. As três primeiras curvas, de cima para baixo, representam os percentis 99, 90 e 50. As duas últimas curvas (linhas tracejadas), os percentis 10 e 1. Pacientes com comprimento telomérico abaixo do percentil 10 ou 1 são considerados curtos ou muito curtos para a idade, respectivamente.

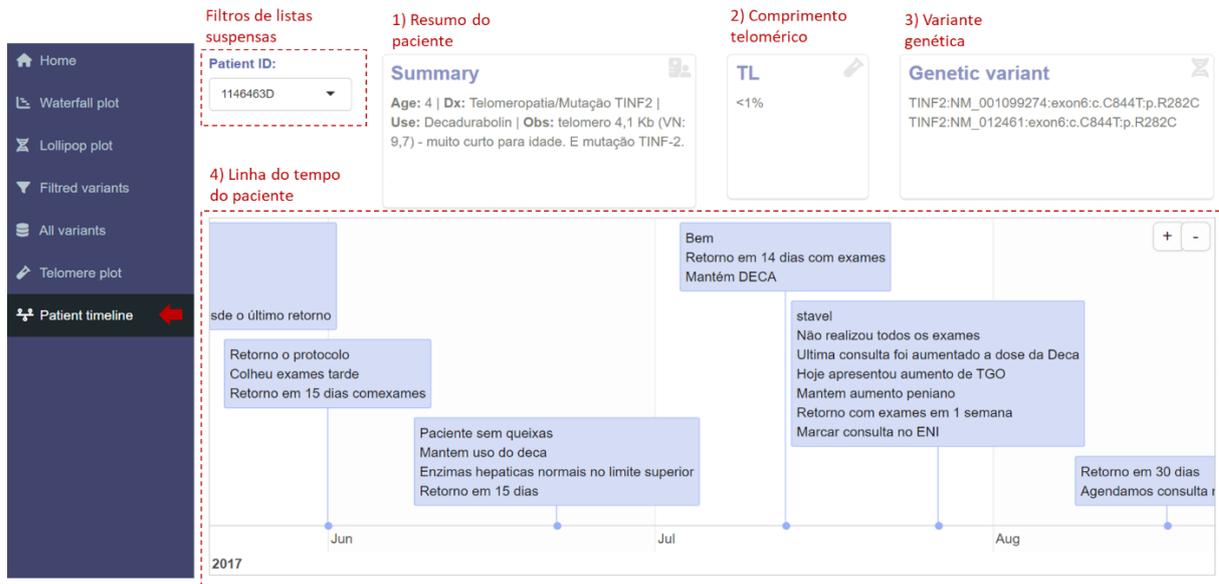
Figura 18 - Página *telomere plot* do aplicativo web *BMF DataViz*.



Fonte: Resultados originais da pesquisa.

Na página *patient timeline*, o paciente pode ser filtrado pelo registro hospitalar (Figura 19). São apresentados quatro quadros contendo, respectivamente, o resumo do paciente, o comprimento telomérico, a(s) variante(s) genética(s) e uma linha do tempo do paciente. No primeiro, tem-se uma descrição resumida do paciente, com idade ao diagnóstico, diagnóstico etiológico ou clínico, medicamento em uso e outras observações. O segundo apresenta o comprimento telomérico. O terceiro quadro mostra uma descrição completa da(s) variante(s) genética(s) associada à clínica. No último temos a linha do tempo com o resumo dos acompanhamentos desse paciente no ambulatório.

Figura 19 - Página *patient timeline* do aplicativo web *BMF DataViz*.



Fonte: Resultados originais da pesquisa.

As análises descritivas apresentadas nos painéis podem ser utilizadas para caracterizar a coorte de pacientes, monitorar a qualidade das informações imputadas no banco, gerar *insights* de novos projetos e/ou realizar estudos de viabilidade de ensaios clínicos.

DISCUSSÃO

5 DISCUSSÃO

5.1 Diagnóstico genético de doenças raras e ultrarraras na população brasileira

O presente trabalho coletou, pela primeira vez, dados genéticos de centenas de pacientes brasileiros com suspeita de SFMHs. Das variantes encontradas em nosso estudo, 32% não havia sido descrita em bancos públicos. Esses achados corroboram com um estudo pioneiro que sequenciou o genoma de 1.171 brasileiros idosos hígidos (NASLAVSKY *et al.*, 2022). Nessa população miscigenada, foram descobertas mais de 2 milhões de variantes que não estão reportadas nos grandes bancos públicos de dados genéticos. O estudo reclassificou a patogenicidade de quase quatrocentas variantes em genes associados a doenças mendelianas, demonstrando a utilidade clínica desse tipo de estudo, em países geneticamente pouco conhecido, para investigação de enfermidades genéticas raras.

Pouco mais da metade (50,3%) das variantes priorizadas em nosso estudo foram classificadas como VUS. Apesar de considerarem um desafio, muitos médicos tratam VUS da mesma maneira como manejam os pacientes com variantes patogênicas (KURIAN *et al.*, 2017). A alta taxa de VUS neste trabalho pode se dar por diferenças étnico-raciais. Alguns estudos mostram que pacientes não-brancos tem maior probabilidade de receber resultados de painel de câncer hereditário com uma variante de significado incerto (CASWELL-JIN *et al.*, 2018). Esses achados enfatizam que a solução para VUS se resume a obtenção de mais dados. Estudos funcionais e dados genéticos em nível populacional melhorarão a classificação das variantes e reduzirão as incertezas (HOFFMAN-ANDREWS, 2017).

Cientes de que baixa representatividade de dados genéticos da população brasileira é um fator limitante para o diagnóstico genético de doenças raras, pesquisadores, financiados pelo Ministério da Saúde, vêm desenvolvendo projetos para mudar esse cenário nos próximos anos. Destaco aqui o projeto Genomas Raros que objetiva sequenciar 8 mil amostras de pacientes com doenças raras acompanhados em centros de referência do sistema público (COELHO *et al.*, 2022). Os achados preliminares mostram uma alta taxa de diagnóstico de 35,7% (FÉLIX *et al.*, 2023).

5.2 Abordagens para diagnóstico genético das SFMH

O painel multigênico (169 genes) utilizado no presente trabalho foi desenhado em parceria com o *Hematology Branch, National Heart, Lung, and Blood Institute* (NHLBI) do *National Institutes of Health* (NIH). Este painel foi capaz de identificar variantes (VUS, LP e P) associadas a falência medular em 46% dos pacientes (taxa de diagnóstico genético).

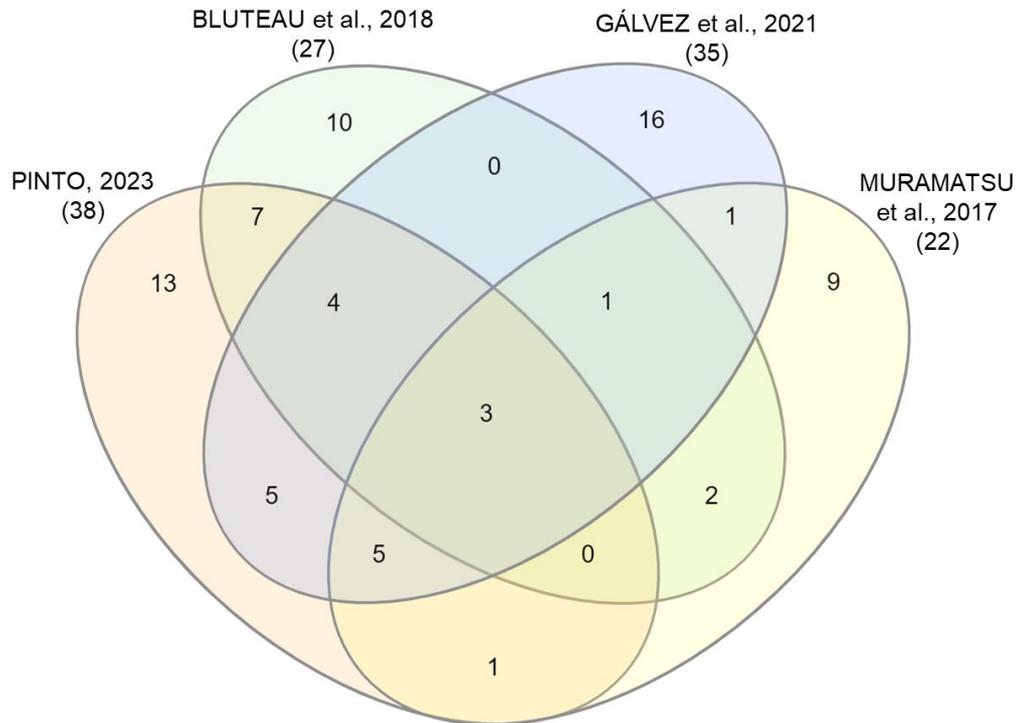
Nos últimos anos, pelo menos, três grandes estudos sequenciaram juntos 754 pacientes com suspeitas de SFMHs e tiveram taxa de diagnóstico genético semelhante ao nosso trabalho. Um grupo japonês utilizou um painel de 178 genes e WES e obtiveram uma taxa de diagnóstico de 44% (MURAMATSU *et al.*, 2017). Um estudo francês encontrou variantes candidatas em 48% dos pacientes, por meio do sequenciamento do exoma total (BLUTEAU *et al.*, 2018). Pesquisadores espanhóis usaram dois painéis, um contendo 129, e outro, 145 genes, e obtiveram uma taxa de diagnóstico genético de 44% (GÁLVEZ *et al.*, 2021). É provável que a estratégia utilizada em nosso trabalho tenha atingido sensibilidade diagnóstica superior a esses três estudos, pois diferente deles, tínhamos como critério de exclusão pacientes com diagnóstico laboratorial de anemia de Fanconi (DEB teste positivo), que poderiam aumentar a porcentagem de casos com alterações genéticas.

Nosso estudo sequenciou 314 pacientes brasileiros com suspeita de falência medular hereditária e encontrou variantes em genes associado as SFMHs em 38 genes. No estudo francês foram avaliados 179 pacientes e 28 genes apresentaram alterações relevantes (BLUTEAU *et al.*, 2018). Dos 371 pacientes japoneses sequenciados no estudo de Muramatsu *et al.*, foram encontradas variantes em 22 genes (MURAMATSU *et al.*, 2017). O estudo espanhol avaliou 204 pacientes e encontrou variantes candidatas em 35 genes (GÁLVEZ *et al.*, 2021).

Comparando o nosso estudo ao espanhol temos em comum 17 genes alterados (*TERC, TERT, DKC1, RTEL1, RPL11, RPL35A, RPL5, RPS26, RPS7, RPS24, RPS19, ELANE, FANCA, FANCG, FANCB, G6PC3* e *MPL*) (Figura 20). Em relação ao francês, encontramos alteração em 14 genes em comum (*TERC, TERT, DKC1, RTEL1, TINF2, PARN, RPL11, RPL35A, RPL5, GATA2, LIG4, MECOM, SAMD9* e *SAMD9L*) (Figura 20). Quando comparamos com o trabalho japonês, esse número cai para 9 genes (*DKC1, TINF2, RPL5, RPS19, RPS24, BRCA2, FANCA, FANCG* e *ELANE*) (Figura 20). Os quatro estudos apresentam em comum apenas três genes

alterados (*DKC1*, *RPL5* e *TINF2*) (Figura 20). Estes resultados exemplificam a heterogeneidade genética das síndromes de falência de medular hereditária entre diferentes populações.

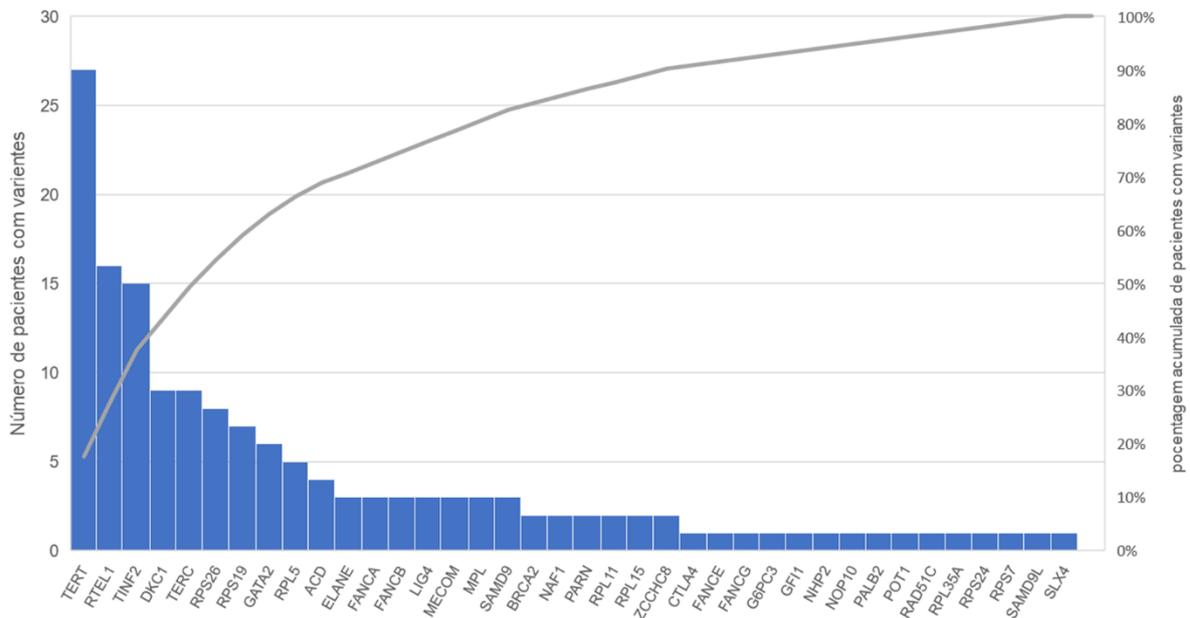
Figura 20 - Diagrama de Venn com os genes alterados em diferentes estudos



Fonte: Resultados originais da pesquisa.

O painel multigênico utilizado no presente trabalho foi desenhado a partir de dados disponíveis em estudos americanos e europeus. Hoje, entendemos que esse painel é bastante amplo e que poderia ser reduzido. A análise de Pareto mostra que 80% dos pacientes que tiverem o diagnóstico genético apresentavam alteração em apenas 16 genes (9,4% do painel) (Figura 21). Painéis grandes para um pequeno grupo de doenças ultrarraras tem custo elevado, gera grande quantidade de dados irrelevantes, dificultando o uso numa rotina clínica. A redução do número dos alvos desse painel diminuiria custos, facilitaria análise e encurtaria o tempo de resposta para o paciente (TAT, do inglês *turnaround time*).

Figura 21 - Gráfico de Pareto dos genes com alterações relevantes pelo número de pacientes.



Fonte: Resultados originais da pesquisa.

Outras abordagens vêm surgindo para auxiliar os clínicos na tomada de decisão na falta do diagnóstico genético. Recentemente, publicamos, em parceria com pesquisadores do NHLBI-NIH um algoritmo de aprendizado de máquina (um subtipo de inteligência artificial) capaz de auxiliar no diagnóstico diferencial de falências medulares, na ausência de testes genéticos (GUTIERREZ-RODRIGUES *et al.*, 2023). O algoritmo utiliza 25 variáveis (laboratoriais e clínicas) para prever a probabilidade de o paciente ter falência medular hereditária, com a acurácia de 92%. Na prática clínica, o algoritmo pode dar suporte a decisões terapêuticas, evitando tratamentos inadequados a espera do diagnóstico genético (GUTIERREZ-RODRIGUES *et al.*, 2023).

No entanto, no contexto das doenças ultrarraras, devemos usar os algoritmos de aprendizado de máquina (AM) com bastante parcimônia e criticidade. Algoritmos de AM treinados em conjuntos pequenos de amostras tendem a apresentar desempenho instável (SCHAEFER *et al.*, 2020). Conjuntos de dados de doenças raras geralmente apresentam bastante ruído nos rótulos das amostras, afetado negativamente os modelos e/ou restringindo o seu uso ao contexto em que as amostras foram treinadas (sobreajuste) (BANERJEE *et al.*, 2023).

5.3 Como a gestão e compartilhamento de dados pode promover o avanço do conhecimento de doenças raras

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) vem incentivando, há alguns anos, o compartilhamento de dados de pesquisas financiadas pela agência. Desde 2017, todos os projetos submetidos à FAPESP devem contar com um Plano de Gestão de Dados (PGD), objetivando tornar públicos os bancos de dados gerados pelos pesquisadores (FAPESP, 2017).

Apesar dessa iniciativa, a cultura da gestão de dados na academia ainda é pouco difundida. Em laboratórios biomédicos, é comum vermos pesquisadores que, ao concluírem seus trabalhos, disponibilizam os dados de pacientes em planilhas digitais, sem controle de acesso, sem garantia de integridade e sem padrão. Essa prática corriqueira fere não apenas os princípios FAIR (do inglês, *findable, accessible, interoperable, and reusable*), como também não está de acordo com o tratamento de dados sensíveis da LGPD (BRASIL, 2018; WILKINSON *et al.*, 2016).

No contexto das doenças raras (DRs), a falta de compartilhamento e padronização dos dados é ainda mais prejudicial, pois o número reduzido de pacientes dificulta conclusões significativas de estudos, diminuindo a chance de desenvolvimento de novos tratamentos (DOS SANTOS VIEIRA *et al.*, 2022; MITANI; HANEUSE, 2020). Quando o tamanho da amostra é pequeno, os famigerados valores de *P* ficam vulneráveis a mínimas variações/desvios. Por falta de pacientes, cerca de um terço de ensaios clínicos com DRs foram descontinuados nos últimos anos (REES *et al.*, 2019). Isso dificulta a entrada de novas terapias de doenças consideradas órfãs.

Iniciativas internacionais tentam solucionar o problema da escassez e da fragmentação dos dados de DRs, com padronização e compartilhamento. A *Rare Diseases Global Open FAIR Implementation* (RDs GO FAIR) objetiva acelerar mudanças culturais entre os pesquisadores que trabalham com DRs e discutir o estabelecimento de padrões e ferramentas que auxiliem na implementação dos princípios FAIR (RDS GO FAIR, 2019).

O consórcio Solve-RD baseou-se nos princípios FAIR para criar a maior rede de compartilhamento de dados de doenças raras. Apoiado pela União Europeia, o Solve-DR reúne mais de 300 médicos, cientistas e representantes de pacientes de 51 centros de 15 países (ZUREK *et al.*, 2021). A principal missão do consórcio é encontrar a causa molecular de diversas DRs (sem diagnóstico) por meio do compartilhamento massivo de dados genômicos e fenotípicos (ZUREK *et al.*, 2021).

O Solve-DR pode ser exemplo para o mundo de como a partilha de dados e de conhecimentos especializados pode promover avanços na compreensão sobre a etiologia de várias doenças raras.

5.4 Aplicação web interativa de visualização de dados genômicos, laboratoriais e clínicos de pacientes com SFMHs

A forma como visualizamos dados científicos praticamente não mudou nas últimas décadas. É comum vermos dados biomédicos (com alto custo de coleta) serem apresentados com métodos inadequados ou pouco eficientes (O'DONOGHUE *et al.*, 2018). A plataforma *BMF DataViz* é um exemplo de solução moderna para melhorar o acesso, a visualização e a maneira como interpretamos dados genômicos, laboratórios e clínicos nos dias atuais.

A ideia da visualização de dados não é impressionar o leitor com imagens complexas. Em geral, simples gráficos de dispersão, de linhas, de barras, histogramas, *boxplots* e mapas de calor resolvem a maioria das necessidades de visualização de dados, quando o objetivo é resumir e auxiliar na identificação de padrões ou tendências (WU *et al.*, 2022).

Atualmente, inexistem plataformas de acesso público para visualização de dados genômicos, laboratoriais ou clínicos de pacientes com SFMHs. No Brasil não temos nenhum repositório multicêntrico de SFMHs. A plataforma *BMF DataViz* visa preencher essa lacuna, dando acesso, inicialmente, a pesquisadores e colaboradores do nosso grupo de pesquisa (HC-FMRP-USP). Uma vez que os dados forem publicados em periódicos, estes poderão ser disponibilizados numa futura versão pública da plataforma.

No que se refere a interatividade, todos os painéis da plataforma *BMF DataViz* contêm listas suspensas para filtro, botões para *download* e/ou gráficos responsivos. Assim, o usuário pode personalizar sua visualização ou baixar arquivos para responder outras perguntas. Pesquisadores discutem que esse tipo de abordagem pode melhorar, inclusive, a reprodutibilidade de estudos (PERKEL, 2018).

CONCLUSÕES E SUGESTÕES
PARA TRABALHOS FUTUROS

6 CONCLUSÕES

- A ferramenta diagnóstica utilizada trouxe resultados conclusivos para quase metade dos pacientes e permitiu a descobertas de dezenas de novas variantes, ampliando o conhecimento genético das SFMHs e mostrando a importância da realização de mais estudos com doenças ultrarraras em população miscigenada e sub-representada em bancos de dados genômicos;
- A utilização do modelo aqui proposto tornará os dados gerados pelo nosso grupo de pesquisa mais localizáveis, acessíveis, interoperáveis e reutilizáveis, criando uma fonte única de dados, facilitando o controle de qualidade e permitindo o armazenamento escalável;
- A plataforma *BMF DataViz* dar acesso rápido a dados genômicos, laboratoriais e clínicos dos pacientes com SFMHs gerados em anos de pesquisas, possibilitando reinterpretação, facilitando a comunicação e geração de *insights* para novos projetos.

7 SUGESTÕES PARA TRABALHOS FUTUROS

- Realizar testes funcionais nas variantes não-descritas;
- Estabelecer um conjunto mínimos de dados para as SFMHs;
- Criar um repositório nacional de dados clínicos, laboratoriais e genômicos de pacientes com SFMHs.

REFERÊNCIAS

8 REFERÊNCIAS

- ALTMANN, T.; GENNERY, A. R. DNA ligase IV syndrome; a review. **Orphanet Journal of Rare Diseases**, v. 11, n. 1, p. 1–7, 7 out. 2016. Disponível em: <<https://ojrd.biomedcentral.com/articles/10.1186/s13023-016-0520-1>>. Acesso em: 27 out. 2023.
- ATTALI, D.; ALMENDE B.V. **timevis: Create Interactive Timeline Visualizations in R2022**. Disponível em: <<https://CRAN.R-project.org/package=timevis>>.
- AUERBACH, A. D. Fanconi anemia and its diagnosis. **Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis**, v. 668, n. 1–2, p. 4–10, 31 jul. 2009. . Acesso em: 21 out. 2023.
- BACIGALUPO, A.; HOWS, J.; GLUCKMAN, E.; NISSEN, C.; MARSH, J.; VAN LINT, M. T.; CONGIU, M.; DE PLANQUE, M. M.; ERNST, P.; MCCANN, S.; RAGAVASHAR, A.; FRICKHOFEN, N.; WURSCH, A.; MARMONT, A. M.; CORDON-SMITH, E. C. Bone marrow transplantation (BMT) versus immunosuppression for the treatment of severe aplastic anaemia (SAA): a report of the EBMT* SAA Working Party. **British Journal of Haematology**, v. 70, n. 2, p. 177–182, 1 out. 1988. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2141.1988.tb02460.x>>. Acesso em: 21 out. 2023.
- BALLMAIER, M.; GERMESHAUSEN, M.; SCHULZE, H.; CHERKAOUI, K.; LANG, S.; GAUDIG, A.; KRUKEMEIER, S.; EILERS, M.; STRAUSS, G.; WELTE, K. c-mpl mutations are the cause of congenital amegakaryocytic thrombocytopenia. **Blood**, v. 97, n. 1, p. 139–146, 1 jan. 2001. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/11133753/>>. Acesso em: 21 out. 2023.
- BANERJEE, J.; TARONI, J. N.; ALLAWAY, R. J.; PRASAD, D. V.; GUINNEY, J.; GREENE, C. Machine learning in rare disease. **Nature Methods** **2023** **20:6**, v. 20, n. 6, p. 803–814, 29 maio 2023. Disponível em: <<https://www.nature.com/articles/s41592-023-01886-z>>. Acesso em: 3 nov. 2023.
- BLUTEAU, O.; SEBERT, M.; LEBLANC, T.; DE LATOUR, R. P.; QUENTIN, S.; LAINEY, E.; HERNANDEZ, L.; DALLE, J. H.; DE FONTBRUNE, F. S.; LENGLINE, E.; ITZYKSON, R.; CLAPPIER, E.; BOISSEL, N.; VASQUEZ, N.; COSTA, M. Da; MASLIAH-PLANCHON, J.; CUCCUINI, W.; RAIMBAULT, A.; DE JAEGERE, L.; ADÈS, L.; FENAUX, P.; MAURY, S.; SCHMITT, C.; MULLER, M.; DOMENECH, C.; BLIN, N.; BRUNO, B.; PELLIER, I.; HUNAUULT, M.; BLANCHE, S.; PETIT, A.; LEVERGER, G.; MICHEL, G.; BERTRAND, Y.; BARUCHEL, A.; SOCIÉ, G.; SOULIER, J. A landscape of germ line mutations in a cohort of inherited bone marrow failure patients. **Blood**, v. 131, n. 7, p. 717–732, 15 fev. 2018. Disponível em: <<https://dx.doi.org/10.1182/blood-2017-09-806489>>. Acesso em: 20 out. 2023.
- BOOCOCK, G. R. B.; MORRISON, J. A.; POPOVIC, M.; RICHARDS, N.; ELLIS, L.; DURIE, P. R.; ROMMENS, J. M. Mutations in SBDS are associated with Shwachman–Diamond syndrome. **Nature Genetics** **2002** **33:1**, v. 33, n. 1, p. 97–

101, 23 dez. 2002. Disponível em: <<https://www.nature.com/articles/ng1062z>>. Acesso em: 21 out. 2023.

BRASIL. **PORTARIA Nº 199, DE 30 DE JANEIRO DE 2014**. Disponível em: <https://bvsms.saude.gov.br/bvs/saudelegis/gm/2014/prt0199_30_01_2014.html>. Acesso em: 4 nov. 2023.

BRASIL. **RESOLUÇÃO Nº 563, DE 10 DE NOVEMBRO DE 2017** <https://conselho.saude.gov.br/resolucoes/2017/Reso563.pdf> 10 nov. 2017.

BRASIL. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>. Acesso em: 4 nov. 2023.

CALADO, R. T.; YOUNG, N. S. Telomere Diseases. <https://doi.org/10.1056/NEJMra0903373>, v. 361, n. 24, p. 2353–2365, 10 dez. 2009. Disponível em: <<https://www.nejm.org/doi/full/10.1056/NEJMra0903373>>. Acesso em: 21 out. 2023.

CASWELL-JIN, J. L.; GUPTA, T.; HALL, E.; PETROVCHICH, I. M.; MILLS, M. A.; KINGHAM, K. E.; KOFF, R.; CHUN, N. M.; LEVONIAN, P.; LEBENSOHN, A. P.; FORD, J. M.; KURIAN, A. W. Racial/ethnic differences in multiple-gene sequencing results for hereditary cancer risk. **Genetics in Medicine**, v. 20, n. 2, p. 234–239, 1 fev. 2018. Disponível em: <<http://www.gimjournal.org/article/S1098360021019080/fulltext>>. Acesso em: 28 out. 2023.

CERAMI, E.; GAO, J.; DOGRUSOZ, U.; GROSS, B. E.; SUMER, S. O.; AKSOY, B. A.; JACOBSEN, A.; BYRNE, C. J.; HEUER, M. L.; LARSSON, E.; ANTIPIN, Y.; REVA, B.; GOLDBERG, A. P.; SANDER, C.; SCHULTZ, N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. **Cancer discovery**, v. 2, n. 5, p. 401–404, maio 2012. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/22588877/>>. Acesso em: 3 nov. 2023.

CHANG, W.; BORGES RIBEIRO, B. **shinydashboard: Create Dashboards with “Shiny”** 2021. Disponível em: <<https://CRAN.R-project.org/package=shinydashboard>>.

CHANG, W.; CHENG, J.; ALLAIRE, J. J.; SIEVERT, C.; SCHLOERKE, B.; XIE, Y.; ALLEN, J.; MCPHERSON, J.; DIPERT, A.; BORGES, B. **shiny: Web Application Framework for R** 2023. Disponível em: <<https://shiny.posit.co/>>.

CODD, E. F. A relational model of data for large shared data banks. **Communications of the ACM**, v. 13, n. 6, p. 377–387, jun. 1970.

COELHO, A. V. C.; MASCARO-CORDEIRO, B.; LUCON, D. R.; NÓBREGA, M. S.; REIS, R. de S.; DE ALEXANDRE, R. B.; MOURA, L. M. S.; OLIVEIRA, G. S. de; GUEDES, R. L. M.; CARACIOLO, M. P.; ZURRO, N. B.; CERVATO, M. C.; OLIVEIRA, J. B. The Brazilian Rare Genomes Project: Validation of Whole Genome Sequencing for Rare Diseases Diagnosis. **Frontiers in Molecular Biosciences**, v. 9,

2 maio 2022. Disponível em: </pmc/articles/PMC9108541/>. Acesso em: 21 out. 2023.

COHEN, I. G.; MELLO, M. M. Big Data, Big Tech, and Protecting Patient Privacy. **JAMA**, v. 322, n. 12, p. 1141–1142, 24 set. 2019. Disponível em: <https://jamanetwork.com/journals/jama/fullarticle/2748399>. Acesso em: 21 out. 2023.

CROOKE, S. T. A call to arms against ultra-rare diseases. **Nature Biotechnology** **2021 39:6**, v. 39, n. 6, p. 671–677, 4 jun. 2021. Disponível em: <https://www.nature.com/articles/s41587-021-00945-0>. Acesso em: 21 out. 2023.

DAVIDSSON, J.; PUSCHMANN, A.; TEDGÅRD, U.; BRYDER, D.; NILSSON, L.; CAMMENGA, J. SAMD9 and SAMD9L in inherited predisposition to ataxia, pancytopenia, and myeloid malignancies. **Leukemia**, v. 32, n. 5, p. 1106, 1 maio 2018. Disponível em: </pmc/articles/PMC5940635/>. Acesso em: 27 out. 2023.

DECHERCHI, S.; PEDRINI, E.; MORDENTI, M.; CAVALLI, A.; SANGIORGI, L. Opportunities and Challenges for Machine Learning in Rare Diseases. **Frontiers in Medicine**, v. 8, p. 747612, 5 out. 2021. . Acesso em: 21 out. 2023.

DOKAL, I. Dyskeratosis congenita in all its forms. **British Journal of Haematology**, v. 110, n. 4, p. 768–779, 1 set. 2000. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-2141.2000.02109.x>. Acesso em: 21 out. 2023.

DOKAL, I.; TUMMALA, H.; VULLIAMY, T. Inherited bone marrow failure in the pediatric patient. **Blood**, v. 140, n. 6, p. 556–570, 11 ago. 2022. Disponível em: <https://dx.doi.org/10.1182/blood.2020006481>. Acesso em: 27 out. 2023.

DOKAL, I.; VULLIAMY, T. Inherited bone marrow failure syndromes. **Haematologica**, v. 95, n. 8, p. 1236–1240, 2010. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/20675743/>. Acesso em: 21 out. 2023.

DOS SANTOS VIEIRA, B.; BERNABÉ, C. H.; ZHANG, S.; ABAZA, H.; BENIS, N.; CÁMARA, A.; CORNET, R.; LE CORNEC, C. M. A.; 'T HOEN, P. A. C.; SCHAEFER, F.; VAN DER VELDE, K. J.; SWERTZ, M. A.; WILKINSON, M. D.; JACOBSEN, A.; ROOS, M. Towards FAIRification of sensitive and fragmented rare disease patient data: challenges and solutions in European reference network registries. **Orphanet Journal of Rare Diseases**, v. 17, n. 1, p. 1–12, 1 dez. 2022. Disponível em: <https://ojrd.biomedcentral.com/articles/10.1186/s13023-022-02558-5>. Acesso em: 27 out. 2023.

DUNHAM, I.; KUNDAJE, A.; ALDRED, S. F.; COLLINS, P. J.; DAVIS, C. A.; DOYLE, F.; EPSTEIN, C. B.; FRIETZE, S.; HARROW, J.; KAUL, R.; KHATUN, J.; LAJOIE, B. R.; LANDT, S. G.; LEE, B. K.; PAULI, F.; ROSENBLOOM, K. R.; SABO, P.; SAFI, A.; SANYAL, A.; SHORESH, N.; SIMON, J. M.; SONG, L.; TRINKLEIN, N. D.; ALTSHULER, R. C.; BIRNEY, E.; BROWN, J. B.; CHENG, C.; DJEBALI, S.; DONG, X.; ERNST, J.; FUREY, T. S.; GERSTEIN, M.; GIARDINE, B.; GREVEN, M.; HARDISON, R. C.; HARRIS, R. S.; HERRERO, J.; HOFFMAN, M. M.; IYER, S.; KELLIS, M.; KHERADPOUR, P.; LASSMANN, T.; LI, Q.; LIN, X.; MARINOV, G. K.;

MERKEL, A.; MORTAZAVI, A.; PARKER, S. C. J.; REDDY, T. E.; ROZOWSKY, J.; SCHLESINGER, F.; THURMAN, R. E.; WANG, J.; WARD, L. D.; WHITFIELD, T. W.; WILDER, S. P.; WU, W.; XI, H. S.; YIP, K. Y.; ZHUANG, J.; BERNSTEIN, B. E.; GREEN, E. D.; GUNTER, C.; SNYDER, M.; PAZIN, M. J.; LOWDON, R. F.; DILLON, L. A. L.; ADAMS, L. B.; KELLY, C. J.; ZHANG, J.; WEXLER, J. R.; GOOD, P. J.; FEINGOLD, E. A.; CRAWFORD, G. E.; DEKKER, J.; ELNITSKI, L.; FARNHAM, P. J.; GIDDINGS, M. C.; GINGERAS, T. R.; GUIGÓ, R.; HUBBARD, T. J.; KENT, W. J.; LIEB, J. D.; MARGULIES, E. H.; MYERS, R. M.; STAMATOYANNOPOULOS, J. A.; TENENBAUM, S. A.; WENG, Z.; WHITE, K. P.; WOLD, B.; YU, Y.; WROBEL, J.; RISK, B. A.; GUNAWARDENA, H. P.; KUIPER, H. C.; MAIER, C. W.; XIE, L.; CHEN, X.; MIKKELSEN, T. S.; GILLESPIE, S.; GOREN, A.; RAM, O.; ZHANG, X.; WANG, L.; ISSNER, R.; COYNE, M. J.; DURHAM, T.; KU, M.; TRUONG, T.; EATON, M. L.; DOBIN, A.; TANZER, A.; LAGARDE, J.; LIN, W.; XUE, C.; WILLIAMS, B. A.; ZALESKI, C.; RÖDER, M.; KOKOCINSKI, F.; ABDELHAMID, R. F.; ALIOTO, T.; ANTOSHECHKIN, I.; BAER, M. T.; BATUT, P.; BELL, I.; BELL, K.; CHAKRABORTTY, S.; CHRAST, J.; CURADO, J.; DERRIEN, T.; DRENKOW, J.; DUMAIS, E.; DUMAIS, J.; DUTTAGUPTA, R.; FASTUCA, M.; FEJES-TOTH, K.; FERREIRA, P.; FOISSAC, S.; FULLWOOD, M. J.; GAO, H.; GONZALEZ, D.; GORDON, A.; HOWALD, C.; JHA, S.; JOHNSON, R.; KAPRANOV, P.; KING, B.; KINGSWOOD, C.; LI, G.; LUO, O. J.; PARK, E.; PREALL, J. B.; PRESAUD, K.; RIBECA, P.; ROBYR, D.; RUAN, X.; SAMMETH, M.; SANDHU, K. S.; SCHAEFFER, L.; SEE, L. H.; SHAHAB, A.; SKANCKE, J.; SUZUKI, A. M.; TAKAHASHI, H.; TILGNER, H.; TROUT, D.; WALTERS, N.; WANG, H.; HAYASHIZAKI, Y.; REYMOND, A.; ANTONARAKIS, S. E.; HANNON, G. J.; RUAN, Y.; CARNINCI, P.; SLOAN, C. A.; LEARNED, K.; MALLADI, V. S.; WONG, M. C.; BARBER, G. P.; CLINE, M. S.; DRESZER, T. R.; HEITNER, S. G.; KAROLCHIK, D.; KIRKUP, V. M.; MEYER, L. R.; LONG, J. C.; MADDREN, M.; RANEY, B. J.; GRASFEDER, L. L.; GIRESI, P. G.; BATTENHOUSE, A.; SHEFFIELD, N. C.; SHOWERS, K. A.; LONDON, D.; BHINGE, A. A.; SHESTAK, C.; SCHANER, M. R.; KIM, S. K.; ZHANG, Z. Z.; MIECZKOWSKI, P. A.; MIECZKOWSKA, J. O.; LIU, Z.; MCDANIELL, R. M.; NI, Y.; RASHID, N. U.; KIM, M. J.; ADAR, S.; ZHANG, Z.; WANG, T.; WINTER, D.; KEEFE, D.; IYER, V. R.; ZHENG, M.; WANG, P.; GERTZ, J.; VIELMETTER, J.; PARTRIDGE, E. C.; VARLEY, K. E.; GASPER, C.; BANSAL, A.; PEPKE, S.; JAIN, P.; AMRHEIN, H.; BOWLING, K. M.; ANAYA, M.; CROSS, M. K.; MURATET, M. A.; NEWBERRY, K. M.; MCCUE, K.; NESMITH, A. S.; FISHER-AYLOR, K. I.; PUSEY, B.; DESALVO, G.; PARKER, S. L.; BALASUBRAMANIAN, S.; DAVIS, N. S.; MEADOWS, S. K.; EGGLESTON, T.; NEWBERRY, J. S.; LEVY, S. E.; ABSHER, D. M.; WONG, W. H.; BLOW, M. J.; VISEL, A.; PENNACHIO, L. A.; PETRYKOWSKA, H. M.; ABYZOV, A.; AKEN, B.; BARRELL, D.; BARSON, G.; BERRY, A.; BIGNELL, A.; BOYCHENKO, V.; BUSSOTTI, G.; DAVIDSON, C.; DESPACIO-REYES, G.; DIEKHANS, M.; EZKURDIA, I.; FRANKISH, A.; GILBERT, J.; GONZALEZ, J. M.; GRIFFITHS, E.; HARTE, R.; HENDRIX, D. A.; HUNT, T.; JUNGREIS, I.; KAY, M.; KHURANA, E.; LENG, J.; LIN, M. F.; LOVELAND, J.; LU, Z.; MANTHRAVADI, D.; MARIOTTI, M.; MUDGE, J.; MUKHERJEE, G.; NOTREDAME, C.; PEI, B.; RODRIGUEZ, J. M.; SAUNDERS, G.; SBONER, A.; SEARLE, S.; SISU, C.; SNOW, C.; STEWARD, C.; TAPANARI, E.; TRESS, M. L.; VAN BAREN, M. J.; WASHIETL, S.; WILMING, L.; ZADISSA, A.; ZHANG, Z.; BRENT, M.; HAUSSLER, D.;

VALENCIA, A.; ADDLEMAN, N.; ALEXANDER, R. P.; AUERBACH, R. K.; BALASUBRAMANIAN, S.; BETTINGER, K.; BHARDWAJ, N.; BOYLE, A. P.; CAO, A. R.; CAYTING, P.; CHAROS, A.; CHENG, Y.; EASTMAN, C.; EUSKIRCHEN, G.; FLEMING, J. D.; GRUBERT, F.; HABEGGER, L.; HARIHARAN, M.; HARMANCI, A.; IYENGAR, S.; JIN, V. X.; KARCZEWSKI, K. J.; KASOWSKI, M.; LACROUTE, P.; LAM, H.; LAMARRE-VINCENT, N.; LIAN, J.; LINDAHL-ALLEN, M.; MIN, R.; MIOTTO, B.; MONAHAN, H.; MOQTADERI, Z.; MU, X. J.; O'GEEN, H.; OUYANG, Z.; PATACSIL, D.; RAHA, D.; RAMIREZ, L.; REED, B.; SHI, M.; SLIFER, T.; WITT, H.; WU, L.; XU, X.; YAN, K. K.; YANG, X.; STRUHL, K.; WEISSMAN, S. M.; PENALVA, L. O.; KARMAKAR, S.; BHANVADIA, R. R.; CHOUDHURY, A.; DOMANUS, M.; MA, L.; MORAN, J.; VICTORSEN, A.; AUER, T.; CENTANIN, L.; EICHENLAUB, M.; GRUHL, F.; HEERMANN, S.; HOECKENDORF, B.; INOUE, D.; KELLNER, T.; KIRCHMAIER, S.; MUELLER, C.; REINHARDT, R.; SCHERTEL, L.; SCHNEIDER, S.; SINN, R.; WITTBRODT, B.; WITTBRODT, J.; JAIN, G.; BALASUNDARAM, G.; BATES, D. L.; BYRON, R.; CANFIELD, T. K.; DIEGEL, M. J.; DUNN, D.; EBERSOL, A. K.; FRUM, T.; GARG, K.; GIST, E.; HANSEN, R. S.; BOATMAN, L.; HAUGEN, E.; HUMBERT, R.; JOHNSON, A. K.; JOHNSON, E. M.; KUTYAVIN, T. V.; LEE, K.; LOTAKIS, D.; MAURANO, M. T.; NEPH, S. J.; NERI, F. V.; NGUYEN, E. D.; QU, H.; REYNOLDS, A. P.; ROACH, V.; RYNES, E.; SANCHEZ, M. E.; SANDSTROM, R. S.; SHAFER, A. O.; STERGACHIS, A. B.; THOMAS, S.; VERNOT, B.; VIERSTRA, J.; VONG, S.; WANG, H.; WEAVER, M. A.; YAN, Y.; ZHANG, M.; AKEY, J. M.; BENDER, M.; DORSCHNER, M. O.; GROUDINE, M.; MACCOSS, M. J.; NAVAS, P.; STAMATOYANNOPOULOS, G.; BEAL, K.; BRAZMA, A.; FLICEK, P.; JOHNSON, N.; LUKK, M.; LUSCOMBE, N. M.; SOBRAL, D.; VAQUERIZAS, J. M.; BATZOGLOU, S.; SIDOW, A.; HUSSAMI, N.; KYRIAZOPOULOU-PANAGIOTOPOULOU, S.; LIBBRECHT, M. W.; SCHAUB, M. A.; MILLER, W.; BICKEL, P. J.; BANFAI, B.; BOLEY, N. P.; HUANG, H.; LI, J. J.; NOBLE, W. S.; BILMES, J. A.; BUSKE, O. J.; SAHU, A. D.; KHARCHENKO, P. V.; PARK, P. J.; BAKER, D.; TAYLOR, J.; LOCHOVSKY, L. An integrated encyclopedia of DNA elements in the human genome. **Nature** 2012 **489**:7414, v. 489, n. 7414, p. 57–74, 5 set. 2012. Disponível em: <<https://www.nature.com/articles/nature11247>>. Acesso em: 4 nov. 2023.

ELMASRI, R.; NAVATHE, S. **Sistemas de Banco de Dados**. 7. ed. São Paulo: Pearson Universidades, 2019. 0–1152 p.

ESMER, C.; SÁNCHEZ, S.; RAMOS, S.; MOLINA, B.; FRIAS, S.; CARNEVALE, A. DEB test for Fanconi anemia detection in patients with atypical phenotypes. **American Journal of Medical Genetics Part A**, v. 124A, n. 1, p. 35–39, 1 jan. 2004. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1002/ajmg.a.20327>>. Acesso em: 21 out. 2023.

FALBEL, D. **auth0: Authentication in Shiny with Auth0**2023. Disponível em: <<https://CRAN.R-project.org/package=auth0>>.

FAPESP. **Gestão de dados**. Disponível em: <<https://fapesp.br/gestaodedados>>. Acesso em: 4 nov. 2023.

FÉLIX, T. M.; FISCHINGER MOURA DE SOUZA, C.; OLIVEIRA, J. B.; RICO-RESTREPO, M.; ZANOTELI, E.; ZATZ, M.; GIUGLIANI, R. Challenges and recommendations to increasing the use of exome sequencing and whole genome sequencing for diagnosing rare diseases in Brazil: an expert perspective.

International Journal for Equity in Health, v. 22, n. 1, p. 1–10, 1 dez. 2023.

Disponível em: <<https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-022-01809-y>>. Acesso em: 21 out. 2023.

FIRKE, S. **janitor: Simple Tools for Examining and Cleaning Dirty Data**2023.

GAGOLEWSKI, M. **stringi: Fast and portable character string processing in R.**

Journal of Statistical Software, v. 103, n. 2, p. 1–59, 2022.

GÁLVEZ, E.; VALLESPÍN, E.; ARIAS-SALGADO, E. G.; SÁNCHEZ-VALDEPEÑAS, C.; GIMÉNEZ, Y.; NAVARRO, S.; RÍO, P.; BOGLIOLO, M.; PUJOL, R.; PEIRÓ, M.; NEVADO, J.; ZUBICARAY, J.; SEBASTIÁN, E.; CATALÁ, A.; BELÉNDEZ, C.; DÍAZ DE HEREDIA, C.; GALERA, A.; BADELL, I.; MADERO, L.; PERONA, R.; SASTRE, L.; SURRALLÉS, J.; BUEREN, J.; LAPUNZINA, P.; SEVILLA, J. Next-generation Sequencing in Bone Marrow Failure Syndromes and Isolated Cytopenias: Experience of the Spanish Network on Bone Marrow Failure Syndromes.

HemaSphere, v. 5, n. 4, 9 abr. 2021. Disponível em:

<https://journals.lww.com/hemasphere/fulltext/2021/04000/next_generation_sequencing_in_bone_marrow_failure.3.aspx>. Acesso em: 4 nov. 2023.

GAO, J.; AKSOY, B. A.; DOGRUSOZ, U.; DRESDNER, G.; GROSS, B.; SUMER, S. O.; SUN, Y.; JACOBSEN, A.; SINHA, R.; LARSSON, E.; CERAMI, E.; SANDER, C.; SCHULTZ, N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. **Science Signaling**, v. 6, n. 269, p. 1–2, 2 abr. 2013. Disponível em: <<https://www.science.org/doi/10.1126/scisignal.2004088>>. Acesso em: 3 nov. 2023.

GERMESHUSEN, M.; ANCLIFF, P.; ESTRADA, J.; METZLER, M.; PONSTINGL, E.; RÜTSCHLE, H.; SCHWABE, D.; SCOTT, R. H.; UNAL, S.; WAWER, A.; ZELLER, B.; BALLMAIER, M. MECOM-associated syndrome: a heterogeneous inherited bone marrow failure syndrome with amegakaryocytic thrombocytopenia. **Blood Advances**, v. 2, n. 6, p. 586, 3 mar. 2018. Disponível em: <[pmc/articles/PMC5873238/](https://pubmed.ncbi.nlm.nih.gov/3004088/)>. Acesso em: 27 out. 2023.

GRANJON, D. **shinydashboardPlus: Add More “AdminLTE2” Components to “shinydashboard”**2021. Disponível em: <<https://CRAN.R-project.org/package=shinydashboardPlus>>.

GRAY, J. The transaction concept: virtues and limitations. **Tandem Computers Incorporated**, v. 81, p. 144–154, 1981.

GUTIERREZ-RODRIGUES, F.; MUNGER, E.; MA, X.; GROARKE, E. M.; TANG, Y.; PATEL, B. A.; CATTO, L. F. B.; CLÉ, D. V.; NIEWISCH, M. R.; ALVES-PAIVA, R. M.; DONAIRES, F. S.; PINTO, A. L.; BORGES, G.; SANTANA, B. A.; MCREYNOLDS, L. J.; GIRI, N.; ALTINTAS, B.; FAN, X.; SHALHOUB, R.; SIWY, C. M.; DIAMOND, C.; RAFFO, D. Q.; CRAFT, K.; KAJIGAYA, S.; SUMMERS, R. M.; LIU, P.;

- CUNNINGHAM, L.; HICKSTEIN, D. D.; DUNBAR, C. E.; PASQUINI, R.; DE OLIVEIRA, M. M.; VELLOSO, E. D. R. P.; ALTER, B. P.; SAVAGE, S. A.; BONFIM, C.; WU, C. O.; CALADO, R. T.; YOUNG, N. S. Differential diagnosis of bone marrow failure syndromes guided by machine learning. **Blood**, v. 141, n. 17, p. 2100–2113, 27 abr. 2023. . Acesso em: 20 out. 2023.
- HARRIS, P. A.; TAYLOR, R.; THIELKE, R.; PAYNE, J.; GONZALEZ, N.; CONDE, J. G. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. **Journal of Biomedical Informatics**, v. 42, n. 2, p. 377–381, 1 abr. 2009. . Acesso em: 14 nov. 2023.
- HAUG, C. J.; DRAZEN, J. M. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. **New England Journal of Medicine**, v. 388, n. 13, p. 1201–1208, 30 mar. 2023. Disponível em: <<https://www.nejm.org/doi/full/10.1056/NEJMra2302038>>. Acesso em: 21 out. 2023.
- HOFFMAN-ANDREWS, L. The known unknown: the challenges of genetic variants of uncertain significance in clinical practice. **Journal of Law and the Biosciences**, v. 4, n. 3, p. 648–657, 1 dez. 2017. Disponível em: <<https://dx.doi.org/10.1093/jlb/lxx038>>. Acesso em: 28 out. 2023.
- HSU, A. P.; JOHNSON, K. D.; FALCONE, E. L.; SANALKUMAR, R.; SANCHEZ, L.; HICKSTEIN, D. D.; CUELLAR-RODRIGUEZ, J.; LEMIEUX, J. E.; ZERBE, C. S.; BRESNICK, E. H.; HOLLAND, S. M. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. **Blood**, v. 121, n. 19, p. 3830–3837, 9 maio 2013. Disponível em: <<https://dx.doi.org/10.1182/blood-2012-08-452763>>. Acesso em: 27 out. 2023.
- INC., P. T. **Collaborative data science** Montreal, QC. Plotly Technologies Inc., 2015. Disponível em: <<https://plot.ly>>.
- KOENKER, R. **quantreg: Quantile Regression** 2022. Disponível em: <<https://CRAN.R-project.org/package=quantreg>>.
- KOPANOS, C.; TSIOLKAS, V.; KOURIS, A.; CHAPPLE, C. E.; ALBARCA AGUILERA, M.; MEYER, R.; MASSOURAS, A. VarSome: the human genomic variant search engine. **Bioinformatics**, v. 35, n. 11, p. 1978–1980, 1 jun. 2019. Disponível em: <<https://dx.doi.org/10.1093/bioinformatics/bty897>>. Acesso em: 24 out. 2023.
- KUEHN, H. S.; OUYANG, W.; LO, B.; DEENICK, E. K.; NIEMELA, J. E.; AVERY, D. T.; SCHICKEL, J. N.; TRAN, D. Q.; STODDARD, J.; ZHANG, Y.; FRUCHT, D. M.; DUMITRIU, B.; SCHEINBERG, P.; FOLIO, L. R.; FREIN, C. A.; PRICE, S.; KOH, C.; HELLER, T.; SEROOGY, C. M.; HUTTENLOCHER, A.; RAO, V. K.; SU, H. C.; KLEINER, D.; NOTARANGELO, L. D.; RAMPERTAAP, Y.; OLIVIER, K. N.; MCELWEE, J.; HUGHES, J.; PITTALUGA, S.; OLIVEIRA, J. B.; MEFFRE, E.; FLEISHER, T. A.; HOLLAND, S. M.; LENARDO, M. J.; TANGYE, S. G.; UZEL, G. Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4. **Science (New York, N.Y.)**, v. 345, n. 6204, p. 1623–1627, 26 set. 2014.

Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/25213377/>>. Acesso em: 27 out. 2023.

KURIAN, A. W.; LI, Y.; HAMILTON, A. S.; WARD, K. C.; HAWLEY, S. T.; MORROW, M.; MCLEOD, M. C.; JAGSI, R.; KATZ, S. J. Gaps in Incorporating Germline Genetic Testing Into Treatment Decision-Making for Early-Stage Breast Cancer. **Journal of Clinical Oncology**, v. 35, n. 20, p. 2232, 7 jul. 2017. Disponível em: </pmc/articles/PMC5501363/>. Acesso em: 28 out. 2023.

LIU, Z.; ZHU, L.; ROBERTS, R.; TONG, W. Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We? **Trends in Genetics**, v. 35, n. 11, p. 852–867, 1 nov. 2019. Disponível em: <<http://www.cell.com/article/S0168952519301799/fulltext>>. Acesso em: 20 out. 2023.

MACHADO, F. **Banco de Dados - Projetos e Implementação**. 4. ed. São Paulo - SP: Editora Érica, 2020.

MARSH, J. C. W.; BALL, S. E.; CAVENAGH, J.; DARBYSHIRE, P.; DOKAL, I.; GORDON-SMITH, E. C.; KEIDAN, J.; LAURIE, A.; MARTIN, A.; MERDIECA, J.; KILLICK, S. B.; STEWART, R.; YIN, J. A. L. Guidelines for the diagnosis and management of aplastic anaemia. **British journal of haematology**, v. 147, n. 1, p. 43–70, out. 2009. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/19673883/>>. Acesso em: 21 out. 2023.

MAYAKONDA, A.; LIN, D.; ASSENOV, Y.; PLASS, C.; KOEFFLER, P. H. Maftools: efficient and comprehensive analysis of somatic variants in cancer. **Genome Research**, 2018.

MITANI, A. A.; HANEUSE, S. Small Data Challenges of Studying Rare Diseases. **JAMA Network Open**, v. 3, n. 3, p. e201965–e201965, 2 mar. 2020. Disponível em: <<https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2763223>>. Acesso em: 27 out. 2023.

MURAMATSU, H.; OKUNO, Y.; YOSHIDA, K.; SHIRAISHI, Y.; DOISAKI, S.; NARITA, A.; SAKAGUCHI, H.; KAWASHIMA, N.; WANG, X.; XU, Y.; CHIBA, K.; TANAKA, H.; HAMA, A.; SANADA, M.; TAKAHASHI, Y.; KANNO, H.; YAMAGUCHI, H.; OHGA, S.; MANABE, A.; HARIGAE, H.; KUNISHIMA, S.; ISHII, E.; KOBAYASHI, M.; KOIKE, K.; WATANABE, K.; ITO, E.; TAKATA, M.; YABE, M.; OGAWA, S.; MIYANO, S.; KOJIMA, S. Clinical utility of next-generation sequencing for inherited bone marrow failure syndromes. **Genetics in Medicine**, v. 19, n. 7, p. 796–802, 1 jul. 2017. . Acesso em: 20 out. 2023.

NASLAVSKY, M. S.; SCLAR, M. O.; YAMAMOTO, G. L.; WANG, J. Y. T.; ZVERINOVA, S.; KARP, T.; NUNES, K.; CERONI, J. R. M.; DE CARVALHO, D. L.; DA SILVA SIMÕES, C. E.; BOZOKLIAN, D.; NONAKA, R.; DOS SANTOS BRITO SILVA, N.; DA SILVA SOUZA, A.; DE SOUZA ANDRADE, H.; PASSOS, M. R. S.; CASTRO, C. F. B.; MENDES-JUNIOR, C. T.; MERCURI, R. L. V.; MILLER, T. L. A.; BUZZO, J. L.; REGO, F. O.; ARAÚJO, N. M.; MAGALHÃES, W. C. S.; MINGRONI-NETTO, R. C.; BORDA, V.; GUIO, H.; ROJAS, C. P.; SANCHEZ, C.; CACERES, O.; DEAN, M.; BARRETO, M. L.; LIMA-COSTA, M. F.; HORTA, B. L.; TARAZONA-

SANTOS, E.; MEYER, D.; GALANTE, P. A. F.; GURYEV, V.; CASTELLI, E. C.; DUARTE, Y. A. O.; PASSOS-BUENO, M. R.; ZATZ, M. Whole-genome sequencing of 1,171 elderly admixed individuals from São Paulo, Brazil. **Nature communications**, v. 13, n. 1, 1 dez. 2022. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/35246524/>>. Acesso em: 21 out. 2023.

NGUENGANG WAKAP, S.; LAMBERT, D. M.; OLRV, A.; RODWELL, C.; GUEYDAN, C.; LANNEAU, V.; MURPHY, D.; LE CAM, Y.; RATH, A. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. **European Journal of Human Genetics** 2019 **28:2**, v. 28, n. 2, p. 165–173, 16 set. 2019. Disponível em: <<https://www.nature.com/articles/s41431-019-0508-0>>. Acesso em: 21 out. 2023.

NICE. **Ultra Orphan Drugs [Internet] - PubMed**. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/28230958/>>. Acesso em: 4 nov. 2023.

O'DONOGHUE, S. I.; BALDI, B. F.; CLARK, S. J.; DARLING, A. E.; HOGAN, J. M.; KAUR, S.; MAIER-HEIN, L.; MCCARTHY, D. J.; MOORE, W. J.; STENAU, E.; SWEDLOW, J. R.; VUONG, J.; PROCTER, J. B. Visualization of Biomedical Data. <https://doi.org/10.1146/annurev-biodatasci-080917-013424>, v. 1, n. 1, p. 275–304, 20 jul. 2018. Disponível em: <<https://www.annualreviews.org/doi/abs/10.1146/annurev-biodatasci-080917-013424>>. Acesso em: 3 nov. 2023.

OLA, O.; SEDIG, K. Discourse with Visual Health Data: Design of Human-Data Interaction. **Multimodal Technologies and Interaction** 2018, **Vol. 2, Page 10**, v. 2, n. 1, p. 10, 20 mar. 2018. Disponível em: <<https://www.mdpi.com/2414-4088/2/1/10/htm>>. Acesso em: 9 nov. 2023.

OUELLETTE, A. C.; MATHEW, J.; MANICKARAJ, A. K.; MANASE, G.; ZAHAVICH, L.; WILSON, J.; GEORGE, K.; BENSON, L.; BOWDIN, S.; MITAL, S. Clinical genetic testing in pediatric cardiomyopathy: Is bigger better? **Clinical genetics**, v. 93, n. 1, p. 33–40, 1 jan. 2018. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/28369760/>>. Acesso em: 20 out. 2023.

PERKEL, J. M. Data visualization tools drive interactivity and reproducibility in online publishing. **Nature** 2021 **554:7690**, 30 jan. 2018. Disponível em: <<https://www.nature.com/articles/d41586-018-01322-9>>. Acesso em: 4 nov. 2023.

R CORE TEAM. **R: A Language and Environment for Statistical Computing** Vienna, Austria. 2020. Disponível em: <<https://www.R-project.org/>>.

RAJKOMAR, A.; DEAN, J.; KOHANE, I. Machine Learning in Medicine. **New England Journal of Medicine**, v. 380, n. 14, p. 1347–1358, 4 abr. 2019. . Acesso em: 21 out. 2023.

RAO, S. S. P.; HUANG, S. C.; GLENN ST HILAIRE, B.; ENGREITZ, J. M.; PEREZ, E. M.; KIEFFER-KWON, K. R.; SANBORN, A. L.; JOHNSTONE, S. E.; BASCOM, G. D.; BOCHKOV, I. D.; HUANG, X.; SHAMIM, M. S.; SHIN, J.; TURNER, D.; YE, Z.; OMER, A. D.; ROBINSON, J. T.; SCHLICK, T.; BERNSTEIN, B. E.; CASELLAS, R.; LANDER, E. S.; AIDEN, E. L. Cohesin Loss Eliminates All Loop Domains. **Cell**, v. 171, n. 2, p. 305- 320.e24, 5 out. 2017. . Acesso em: 4 nov. 2023.

RDS GO FAIR. **Rare Diseases - GO FAIR**. Disponível em: <<https://www.go-fair.org/implementation-networks/overview/rare-diseases/>>. Acesso em: 4 nov. 2023.

REES, C. A.; PICA, N.; MONUTEAUX, M. C.; BOURGEOIS, F. T. Noncompletion and nonpublication of trials studying rare diseases: A cross-sectional analysis. **PLOS Medicine**, v. 16, n. 11, p. e1002966, 2019. Disponível em: <<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002966>>. Acesso em: 27 out. 2023.

RICHARDS, S.; AZIZ, N.; BALE, S.; BICK, D.; DAS, S.; GASTIER-FOSTER, J.; GRODY, W. W.; HEGDE, M.; LYON, E.; SPECTOR, E.; VOELKERDING, K.; REHM, H. L. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. **Genetics in Medicine** 2015 **17:5**, v. 17, n. 5, p. 405–423, 5 mar. 2015. Disponível em: <<https://www.nature.com/articles/gim201530>>. Acesso em: 24 out. 2023.

RSTUDIO TEAM. **RStudio: Integrated Development Environment for R** Boston, MA. 2020. Disponível em: <<http://www.rstudio.com/>>.

SCHAEFER, J.; LEHNE, M.; SCHEPERS, J.; PRASSER, F.; THUN, S. The use of machine learning in rare diseases: A scoping review. **Orphanet Journal of Rare Diseases**, v. 15, n. 1, p. 1–10, 9 jun. 2020. Disponível em: <<https://ojrd.biomedcentral.com/articles/10.1186/s13023-020-01424-6>>. Acesso em: 3 nov. 2023.

SHIMAMURA, A.; ALTER, B. P. Pathophysiology and management of inherited bone marrow failure syndromes. **Blood Reviews**, v. 24, n. 3, p. 101–122, 1 maio 2010. . Acesso em: 21 out. 2023.

TOWNSLEY, D. M.; DUMITRIU, B.; YOUNG, N. S. Bone marrow failure and the telomeropathies. **Blood**, v. 124, n. 18, p. 2775–2783, 30 out. 2014. Disponível em: <<https://dx.doi.org/10.1182/blood-2014-05-526285>>. Acesso em: 21 out. 2023.

TSANGARIS, E.; KLAASSEN, R.; FERNANDEZ, C. V.; YANOFSKY, R.; SHERECK, E.; CHAMPAGNE, J.; SILVA, M.; LIPTON, J. H.; BROSSARD, J.; MICHON, B.; ABISH, S.; STEELE, M.; ALI, K.; DOWER, N.; ATHALE, U.; JARDINE, L.; HAND, J. P.; ODAME, I.; CANNING, P.; ALLEN, C.; CARCAO, M.; BEYENE, J.; ROIFMAN, C. M.; DROR, Y. Genetic analysis of inherited bone marrow failure syndromes from one prospective, comprehensive and population-based cohort and identification of novel mutations. **Journal of Medical Genetics**, v. 48, n. 9, p. 618–628, 1 set. 2011. Disponível em: <<https://jmg.bmj.com/content/48/9/618>>. Acesso em: 27 out. 2023.

VISIBELLI, A.; RONCAGLIA, B.; SPIGA, O.; SANTUCCI, A. The Impact of Artificial Intelligence in the Odyssey of Rare Diseases. **Biomedicines** 2023, **Vol. 11, Page 887**, v. 11, n. 3, p. 887, 13 mar. 2023. Disponível em: <<https://www.mdpi.com/2227-9059/11/3/887/htm>>. Acesso em: 21 out. 2023.

VLACHOS, A.; BALL, S.; DAHL, N.; ALTER, B. P.; SHETH, S.; RAMENGI, U.; MEERPOHL, J.; KARLSSON, S.; LIU, J. M.; LEBLANC, T.; PALEY, C.; KANG, E. M.; LEDER, E. J.; ATSIDAFTOS, E.; SHIMAMURA, A.; BESSLER, M.; GLADER, B.;

LIPTON, J. M. Diagnosing and treating Diamond Blackfan anaemia: results of an international clinical consensus conference. **British Journal of Haematology**, v. 142, n. 6, p. 859–876, 1 set. 2008. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2141.2008.07269.x>>. Acesso em: 21 out. 2023.

VULLIAMY, T.; DOKAL, I. Dyskeratosis Congenita. **Seminars in Hematology**, v. 43, n. 3, p. 157–166, 1 jul. 2006. . Acesso em: 21 out. 2023.

WEGMAN-OSTROSKY, T.; SAVAGE, S. A. The genomics of inherited bone marrow failure: from mechanism to the clinic. **British Journal of Haematology**, v. 177, n. 4, p. 526–542, 1 maio 2017. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/bjh.14535>>. Acesso em: 21 out. 2023.

WELTE, K.; ZEIDLER, C. Severe Congenital Neutropenia. **Hematology/Oncology Clinics of North America**, v. 23, n. 2, p. 307–320, 1 abr. 2009. . Acesso em: 21 out. 2023.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. [s.l.] Springer-Verlag New York, 2016.

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K.; VAUGHAN, D. **dplyr: A Grammar of Data Manipulation**2023.

WICKHAM, H.; VAUGHAN, D.; GIRLICH, M. **tidyr: Tidy Messy Data**2023. Disponível em: <<https://tidyr.tidyverse.org>>.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, Ij. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J. W.; DA SILVA SANTOS, L. B.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; T HOEN, P. A. C.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; VAN SCHAİK, R.; SANSONE, S. A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; VAN DER LEI, J.; VAN MULLIGEN, E.; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data** 2016 **3:1**, v. 3, n. 1, p. 1–9, 15 mar. 2016. Disponível em: <<https://www.nature.com/articles/sdata201618>>. Acesso em: 21 out. 2023.

WU, T.; XIAO, C. Le; LAM, T. T. Y.; YU, G. Editorial: Biomedical Data Visualization: Methods and Applications. **Frontiers in Genetics**, v. 13, p. 27, 27 abr. 2022. Disponível em: <<https://pmc/articles/PMC9092592/>>. Acesso em: 3 nov. 2023.

XIE, Y.; CHENG, J.; TAN, X. **DT: A Wrapper of the JavaScript Library “DataTables”**2022. Disponível em: <<https://CRAN.R-project.org/package=DT>>.

YANG, H.; WANG, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. **Nature Protocols** 2015 10:10, v. 10, n. 10, p. 1556–1566, 17 set. 2015. Disponível em: <<https://www.nature.com/articles/nprot.2015.105>>. Acesso em: 22 out. 2023.

YOUNG, N. S. Aplastic Anemia. <https://doi.org/10.1056/NEJMra1413485>, v. 379, n. 17, p. 1643–1656, 24 out. 2018. Disponível em: <<https://www.nejm.org/doi/10.1056/NEJMra1413485>>. Acesso em: 27 out. 2023.

YOUNG, N. S.; CALADO, R. T.; SCHEINBERG, P. Current concepts in the pathophysiology and treatment of aplastic anemia. **Blood**, v. 108, n. 8, p. 2509–2519, 15 out. 2006. Disponível em: <<https://dx.doi.org/10.1182/blood-2006-03-010777>>. Acesso em: 21 out. 2023.

ZUREK, B.; ELLWANGER, K.; VISSERS, L. E. L. M.; SCHÜLE, R.; SYNOFZIK, M.; TÖPF, A.; DE VOER, R. M.; LAURIE, S.; MATALONGA, L.; GILISSEN, C.; OSSOWSKI, S.; 'T HOEN, P. A. C.; VITOBELLO, A.; SCHULZE-HENTRICH, J. M.; RIESS, O.; BRUNNER, H. G.; BROOKES, A. J.; RATH, A.; BONNE, G.; GUMUS, G.; VERLOES, A.; HOOPERBRUGGE, N.; EVANGELISTA, T.; HARMUTH, T.; SWERTZ, M.; SPALDING, D.; HOISCHEN, A.; BELTRAN, S.; GRAESSNER, H.; HAACK, T. B.; ZUREK, B.; ELLWANGER, K.; DEMIDOV, G.; STURM, M.; KESSLER, C.; WAYAND, M.; WILKE, C.; TRASCHÜTZ, A.; SCHÖLS, L.; HENGEL, H.; HEUTINK, P.; BRUNNER, H.; SCHEFFER, H.; STEYAERT, W.; SABLAUSKAS, K.; DE VOER, R. M.; KAMSTEEG, E. J.; VAN DE WARRENBURG, B.; VAN OS, N.; TE PASKE, I.; JANSSEN, E.; DE BOER, E.; STEEHOUWER, M.; YALDIZ, B.; KLEEFSTRA, T.; VEAL, C.; GIBSON, S.; WADSLEY, M.; MEHTARIZADEH, M.; RIAZ, U.; WARREN, G.; DIZJIKAN, F. Y.; SHORTER, T.; STRAUB, V.; BETTOLO, C. M.; SPECHT, S.; CLAYTON-SMITH, J.; BANKA, S.; ALEXANDER, E.; JACKSON, A.; FAIVRE, L.; THAUVIN, C.; VITOBELLO, A.; DENOMMÉ-PICHON, A. S.; DUFFOURD, Y.; TISSERANT, E.; BRUEL, A. L.; PEYRON, C.; PÉLISSIER, A.; BELTRAN, S.; GUT, I. G.; LAURIE, S.; PISCIA, D.; MATALONGA, L.; PAPAKONSTANTINO, A.; BULLICH, G.; CORVO, A.; GARCIA, C.; FERNANDEZ-CALLEJO, M.; HERNÁNDEZ, C.; PICÓ, D.; PARAMONOV, I.; LOCHMÜLLER, H.; GUMUS, G.; BROS-FACER, V.; HANAUER, M.; OLRÝ, A.; LAGORCE, D.; HAVRYLENKO, S.; IZEM, K.; RIGOUR, F.; STEVANIN, G.; DURR, A.; DAVOINE, C. S.; GUILLOT-NOEL, L.; HEINZMANN, A.; COARELLI, G.; ALLAMAND, V.; NELSON, I.; YAOU, R. Ben; METAY, C.; EYMARD, B.; COHEN, E.; ATALAIA, A.; STOJKOVIC, T.; MACEK, M.; TURNOVEC, M.; THOMASOVÁ, D.; KREMLIKOVÁ, R. P.; FRANKOVÁ, V.; HAVLOVICOVÁ, M.; KREMLIK, V.; PARKINSON, H.; KEANE, T.; SENF, A.; ROBINSON, P.; DANIS, D.; ROBERT, G.; COSTA, A.; PATCH, C.; HANNA, M.; HOULDEN, H.; REILLY, M.; VANDROVCOVA, J.; MUNTONI, F.; ZAHARIEVA, I.; SARKOZY, A.; TIMMERMAN, V.; BAETS, J.; VAN DE VONDEL, L.; BEIJER, D.; DE JONGHE, P.; NIGRO, V.; BANFI, S.; TORELLA, A.; MUSACCHIA, F.; PILUSO, G.; FERLINI, A.; SELVATICI, R.; ROSSI, R.; NERI, M.; ARETZ, S.; SPIER, I.; SOMMER, A. K.; PETERS, S.; OLIVEIRA, C.; PELAEZ, J. G.; MATOS, A. R.; JOSÉ, C. S.; FERREIRA, M.; GULLO, I.; FERNANDES, S.; GARRIDO, L.; FERREIRA, P.; CARNEIRO, F.; JOHANSSON, L.; VAN DER VELDE, J. K.; VAN DER VRIES, G.; NEERINCX, P. B.; ROELOFS-PRINS, D.; KÖHLER, S.;

METCALFE, A.; VERLOES, A.; DRUNAT, S.; ROORYCK, C.; TRIMOUILLE, A.; CASTELLO, R.; MORLEO, M.; PINELLI, M.; VARAVALLO, A.; DE LA PAZ, M. P.; SÁNCHEZ, E. B.; MARTÍN, E. L.; DELGADO, B. M.; DE LA ROSA, F. J. A. G.; CIOLFI, A.; DALLAPICCOLA, B.; PIZZI, S.; RADIO, F. C.; TARTAGLIA, M.; RENIERI, A.; BENETTI, E.; BALICZA, P.; MOLNAR, M. J.; MAVER, A.; PETERLIN, B.; MÜNCHAU, A.; LOHMANN, K.; HERZOG, R.; PAULY, M.; MACAYA, A.; MARCÉ-GRAU, A.; OSORIO, A. N.; DE BENITO, D. N.; LOCHMÜLLER, H.; THOMPSON, R.; POLAVARAPU, K.; BEESON, D.; COSSINS, J.; CRUZ, P. M. R.; HACKMAN, P.; JOHARI, M.; SAVARESE, M.; UDD, B.; HORVATH, R.; CAPELLA, G.; VALLE, L.; HOLINSKI-FEDER, E.; LANER, A.; STEINKE-LANGE, V.; SCHRÖCK, E.; RUMP, A. Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases. **European Journal of Human Genetics** 2021 29:9, v. 29, n. 9, p. 1325–1331, 1 jun. 2021. Disponível em: <<https://www.nature.com/articles/s41431-021-00859-0>>. Acesso em: 27 out. 2023.

APÊNDICE A - Genes associados a Síndrome de Falência de Medula Óssea hereditária

SFMO	Gene	padrão de herança
anemia de Fanconi	<i>FANCA</i>	AR
	<i>FANCC</i>	AR
	<i>BRCA2 (FANCD1)</i>	AR
	<i>FANCD2</i>	AR
	<i>FANCE</i>	AR
	<i>FANCF</i>	AR
	<i>FANCG (XRCC9)</i>	AR
	<i>FANCI</i>	AR
	<i>BRIP1 (FANCJ)</i>	AR
	<i>FANCL</i>	AR
	<i>FANCM</i>	AR
	<i>PALB2 (FANCN)</i>	AR
	<i>RAD51C (FANCO)</i>	AR
	<i>SLX4 (FANCP)</i>	AR
	<i>ERCC4 (FANCQ, XPF)</i>	AR
	<i>RAD51 (FANCR)</i>	AR
	<i>BRCA1 (FANCS)</i>	AR
	<i>UBE2T (FANCT)</i>	AR
	<i>XRCC2 (FANCU)</i>	AR
	<i>MAD2L2 (REV7)</i>	AR
<i>FANCB</i>	XLR	
Telomeropatia	<i>DKC1</i>	XLR
	<i>TINF2</i>	AD
	<i>TERT</i>	AD e AR
	<i>TERC</i>	AD
	<i>RTEL1</i>	AD e AR
	<i>ACD</i>	AR
	<i>CTC1</i>	AR
	<i>NOP10</i>	AR
	<i>NHP2</i>	AR
	<i>PARN</i>	AR e AD
	<i>WRAP53</i>	AR
	<i>NAF1</i>	AD
	<i>ZCCHC8</i>	AD
	<i>POT1</i>	AD
<i>TPP1</i>	AR	

SFMO	Gene	padrão de herança
Ribosomopatia	<i>RPS19</i>	AD
	<i>RPL11</i>	AD
	<i>RPS26</i>	AD
	<i>RPS10</i>	AD
	<i>RPL35A</i>	AD
	<i>RPS24</i>	AD
	<i>RPS17</i>	AD
	<i>RPL5</i>	AD
	<i>RPL15</i>	AD
	<i>RPL17</i>	AD
	<i>RPL19</i>	AD
	<i>RPL26</i>	AD
	<i>RPL31</i>	AD
	<i>RPS7</i>	AD
	<i>RPS19</i>	AD
	<i>RPS20</i>	AD
	<i>RPS28</i>	AD
	<i>RPS29</i>	AD
	<i>GATA1</i>	XLR
	<i>SBDS</i>	AR
deficiência GATA2	<i>GATA2</i>	AD
neutropenia congênita grave	<i>ELANE (ELA2)</i>	AD
	<i>HAX1</i>	AR
	<i>G6PC3</i>	AR
	<i>GFI1</i>	AR
	<i>WAS</i>	XLR
	<i>JAGN1</i>	AR
	<i>CSF3R</i>	AR
	<i>TCIRG1</i>	AD
deficiência MECOM	<i>MECOM</i>	AD
deficiência SAMD9	<i>SAMD9L</i>	AD
	<i>SAMD9</i>	AD
deficiência CTLA4	<i>CTLA4</i>	AD
deficiência LIG4	<i>LIG4</i>	AR
trombocitopenia amegacariocítica congênita	<i>MPL</i>	AR

Abreviações: AD (autossômico dominante), AR (autossômico recessivo) e XLR (recessiva ligada ao X)

APÊNDICE B - Tabela de diagnóstico e alterações genéticas dos pacientes com suspeita de Síndrome de Falência de Medula Óssea hereditária

ID	Sexo	Telômero	Diagnóstico genético	Variante	Função gênica	Função exônica	GT	dbSNP	Freq. Máx. pop	Class. ACMG	Freq. alélica	Bancos públicos
P001	M	curto	telomeropatia	DKC1:NM_001142463:exon11:c.C1058T:p.A353V	exônica	não sinônimo	1/1	rs121912288	0	LP	1	sim
P002	M	curto	telomeropatia	TINF2:NM_001099274:exon6:c.G845A:p.R282H	exônica	não sinônimo	0/1	rs121918544		P	0,449	sim
P003	M	curto	telomeropatia	DKC1:NM_001142463:exon4:c.A196G:p.T66A	exônica	não sinônimo	1/1	rs121912297		P	1	sim
P004	M	curto	telomeropatia	TINF2:NM_001099274:exon6:c.851_852del:p.T284Sfs*7	exônica	deleção com frameshift	0/1			LP	0,503	não
P005	M	curto	telomeropatia	TERT:NM_001193376:exon11:c.C2722T:p.R908C	exônica	não sinônimo	0/1			VUS	0,518	sim
P006	F	curto	telomeropatia	RTEL1:NM_001283010:exon31:c.A2588G:p.Y863C	exônica	não sinônimo	0/1		0,0000298	VUS	0,471	sim
P006	F	curto	telomeropatia	RTEL1:NM_001283009:exon34:c.T3730C:p.C1244R	exônica	não sinônimo	0/1	rs587777037	5,809E-05	VUS	0,431	sim
P007	M	normal	telomeropatia	NAF1:NM_001128931:exon3:c.T587G:p.I196S	exônica	não sinônimo	0/1			VUS	0,451	não
P008	F	curto	telomeropatia	TERT:NM_001193376:exon2:c.C1072T:p.R358W	exônica	não sinônimo	1/1			VUS	0,995	sim
P009	F	curto	telomeropatia	TINF2:NM_001099274:exon6:c.G845A:p.R282H	exônica	não sinônimo	0/1	rs121918544		P	0,467	sim
P010	M	curto	telomeropatia	TERT:NM_001193376:exon3:c.1714delA:p.R572Gfs*22	exônica	deleção com frameshift	0/1			LP	0,412	não
P011	M	curto	telomeropatia	TERT:NM_001193376:exon11:c.C2722T:p.R908C	exônica	não sinônimo	0/1			VUS	0,699	sim
P012	M	curto	telomeropatia	TERT:NM_001193376:exon5:c.C2086T:p.R696C	exônica	não sinônimo	1/1			VUS	1	não
P013	M	curto	telomeropatia	TERT:NM_001193376:exon10:c.G2594A:p.R865H	exônica	não sinônimo	0/1	rs121918666	0,0002	LP	0,494	sim
P014	F	curto	telomeropatia	RTEL1:NM_001283009:exon34:c.T3730C:p.C1244R	exônica	não sinônimo	0/1	rs587777037	5,809E-05	VUS	0,5	sim
P015	F	curto	telomeropatia	TERT:NM_001193376:exon10:c.G2594A:p.R865H	exônica	não sinônimo	0/1	rs121918666	0,0002	LP	0,466	sim
P016	M	curto	telomeropatia	DKC1:NM_001142463:exon11:c.C1058T:p.A353V	exônica	não sinônimo	1/1	rs121912288	0	LP	0,993	sim
P017	F	curto	telomeropatia	TINF2:NM_001099274:exon6:c.T869G:p.F290C	exônica	não sinônimo	0/1			LP	0,481	não
P018	F	curto	telomeropatia	TERT:NM_001193376:exon14:c.C3045G:p.F1015L	exônica	não sinônimo	0/1			VUS	0,468	não

P019	M	curto	telomeropatia	TINF2:NM_001099274:exon6:c.G845A:p.R282H	exônica	não sinônimo	0/1	rs121918544		P	0,45	sim
P020	F	curto	telomeropatia	TERT:NM_001193376:exon10:c.G2594A:p.R865H	exônica	não sinônimo	0/1	rs121918666	0,0002	LP	0,453	sim
P021	M	curto	telomeropatia	TERT:NM_001193376:exon6:c.T2218C:p.C740R	exônica	não sinônimo	0/1			VUS	0,461	não
P022	M	curto	telomeropatia	TINF2:NM_001099274:exon6:c.G845A:p.R282H	exônica	não sinônimo	0/1	rs121918544		P	0,5	sim
P023	M	curto	telomeropatia	TINF2:NM_001099274:exon6:c.C844T:p.R282C	exônica	não sinônimo	0/1	rs121918545	0	P	0,469	sim
P024	F	curto	telomeropatia	NHP2:NM_001034833:exon2:c.G190A:p.V64M	exônica	não sinônimo	0/1	rs79031130	0,0011	VUS	0,468	sim
P024	F	curto	telomeropatia	NAF1:NM_138386:exon8:c.A1478C:p.Y493S	exônica	não sinônimo	0/1	rs143001503	0,01	VUS	0,491	sim
P025	F	curto	telomeropatia	DKC1:NM_001142463:exon2:c.40_42del:p.K17del	exônica	deleção sem frameshift	0/1	rs782201995	0,001	VUS	0,383	sim
P026	M	curto	telomeropatia	TERC:NR_001566::r.110g>c	ncRNA		0/1			VUS	0,452	não
P027	F	curto	telomeropatia	TINF2:NM_001099274:exon6:c.C844T:p.R282C	exônica	não sinônimo	0/1	rs121918545	0	P	0,449	sim
P028	M	curto	telomeropatia	RTEL1:NM_001283009:exon34:c.3775_3776del:p.A1259Lfs*2	exônica	deleção com frameshift	0/1			LP	0,548	não
P029	F	curto	telomeropatia	POT1:NM_001042594:exon7:c.C44T:p.P15L	exônica	não sinônimo	0/1			VUS	0,449	sim
P030	M	curto	telomeropatia	RTEL1:NM_001283010:exon31:c.A2588G:p.Y863C	exônica	não sinônimo	0/1		0,0000298	VUS	0,489	sim
P030	M	curto	telomeropatia	RTEL1:NM_001283009:exon34:c.3775_3776del:p.A1259Lfs*2	exônica	deleção com frameshift	0/1			LP	0,44	não
P031	M	curto	telomeropatia	DKC1:NM_001142463:exon11:c.C1058T:p.A353V	exônica	não sinônimo	1/1	rs121912288	0	LP	1	sim
P032	F	curto	telomeropatia	RTEL1:NM_001283010:exon11:c.G302A:p.R101H	exônica	não sinônimo	0/1	rs113684274	0,0011	VUS	0,5	sim
P033	F	curto	telomeropatia	RTEL1:NM_001283010:exon31:c.A2588G:p.Y863C	exônica	não sinônimo	0/1		0,0000298	VUS	0,472	sim
P033	F	curto	telomeropatia	RTEL1:NM_001283009:exon34:c.T3730C:p.C1244R	exônica	não sinônimo	0/1	rs587777037	5,809E-05	VUS	0,5	sim
P034	M	curto	telomeropatia	TINF2:NM_001099274:exon6:c.839delA:p.K280Rfs*37	exônica	deleção com frameshift	0/1			P	0,421	não
P035	M	curto	telomeropatia	DKC1:NM_001142463:exon11:c.C1058T:p.A353V	exônica	não sinônimo	1/1	rs121912288	0	LP	0,939	sim
P036	M	curto	telomeropatia	ZCCHC8:NM_001350935:exon1:c.G100A:p.D34N	exônica	não sinônimo	0/1		1,149E-05	VUS	0,486	sim
P037	M	curto	telomeropatia	TERT:NM_001193376:exon6:c.G2146A:p.A716T	exônica	não sinônimo	0/1	rs387907249		LP	0,461	sim

P038	M	curto	telomeropatia	TERT:NM_001193376:exon9:c.G2578A:p.D860N	exônica	não sinônimo	1/1		VUS	0,989	sim
P039	M	curto	telomeropatia	PARN:NM_001242992:exon11:c.C695T:p.A232V	exônica	não sinônimo	0/1		VUS	0,464	não
P040	M	curto	telomeropatia	TERC:NR_001566::n.94C>T	ncRNA		0/1		VUS	0,452	não
P040	M	curto	telomeropatia	TERT:NM_001193376:exon7:c.G2329A:p.V777M	exônica	não sinônimo	0/1		VUS	0,463	sim
P041	F	curto	telomeropatia	TERC:NR_001566::n.110_113delGACT	ncRNA		0/1	rs199422270	LP	0,592	sim
P042	M	curto	telomeropatia	TERT:NM_001193376:exon6:c.C2154A:p.D718E	exônica	não sinônimo	0/1		VUS	0,466	não
P043	M	curto	telomeropatia	TERT:NM_001193376:exon11:c.C2722T:p.R908C	exônica	não sinônimo	0/1		VUS	0,477	sim
P043	M	curto	telomeropatia	TERT:NM_198253:exon11:c.T2780G:p.L927R	exônica	não sinônimo	0/1		VUS	0,466	não
P044	M	curto	telomeropatia	TERC:NR_001566::n.110_113delGACT	ncRNA		0/1	rs199422270	LP	0,426	sim
P045	M	curto	telomeropatia	DKC1:NM_001142463:exon11:c.C1151A:p.P384Q	exônica	não sinônimo	1/1		VUS	1	não
P046	F	curto	telomeropatia	TERT:NM_001193376:exon3:c.G1730A:p.R577Q	exônica	não sinônimo	0/1	1,792E-05	VUS	0,533	sim
P047	M	curto	telomeropatia	DKC1:NM_001142463:exon3:c.T113C:p.I38T	exônica	não sinônimo	1/1	rs28936072	LP	0,997	sim
P048	F	curto	telomeropatia	ACD:NM_001082486:exon3:c.G535C:p.G179R	exônica	não sinônimo	1/1	9,013E-06	VUS	1	sim
P049	M	curto	telomeropatia	TINF2:NM_001099274:exon6:c.848_849insGCCC:p.T284Pfs*9	exônica	inserção com frameshift	0/1		P	0,451	não
P050	M	curto	telomeropatia	DKC1:NM_001142463:exon1:c.C5T:p.A2V	exônica	não sinônimo	1/1	rs121912303	LP	0,933	sim
P051	M	curto	telomeropatia	TERT:NM_001193376:exon11:c.G2659C:p.A887P	exônica	não sinônimo	0/1		VUS	0,454	não
P051	M	curto	telomeropatia	RTEL1:NM_001283009:exon8:c.T619C:p.C207R	exônica	não sinônimo	0/1		LP	0,517	não
P052	M	curto	telomeropatia	TERT:NM_001193376:exon4:c.G1885A:p.G629R	exônica	não sinônimo	0/1		VUS	0,46	sim
P053	M	curto	telomeropatia	TERT:NM_001193376:exon3:c.G1730A:p.R577Q	exônica	não sinônimo	0/1	1,792E-05	VUS	0,449	sim
P054	F	curto	telomeropatia	TINF2:NM_001099274:exon6:c.G845A:p.R282H	exônica	não sinônimo	0/1	rs121918544	P	0,487	sim
P055	M	curto	telomeropatia	TERC:NR_001566::r.260_261insgcg	ncRNA		0/1		LP	0,45	sim
P056	F	curto	telomeropatia	TERT:NM_001193376:exon4:c.C1891T:p.R631W	exônica	não sinônimo	0/1	8,955E-06	LP	0,477	sim
P057	M	curto	telomeropatia	TERT:NM_001193376:exon2:c.G1492A:p.G498R	exônica	não sinônimo	0/1		VUS	0,56	sim
P057	M	curto	telomeropatia	RTEL1:NM_001283009:exon5:c.403_404del:p.V137Afs*12	exônica	deleção com frameshift	0/1		LP	0,482	não

P058	F	curto	telomeropatia	TINF2:NM_001099274:exon6:c.C844T:p.R282C	exônica	não sinônimo	0/1	rs121918545	0	P	0,504	sim
P059	M	curto	telomeropatia	ACD:NM_001082486:exon2:c.C364T:p.Q122X	exônica	ganho de parada	0/1		0	LP	0,468	sim
P060	M	curto	telomeropatia	ACD:NM_001082486:exon12:c.1622_1625del:p.P541Lfs*35	exônica	deleção com frameshift	0/1	rs772312742	0,0002	VUS	0,3	sim
P061	M	curto	telomeropatia	TERT:NM_001193376:exon3:c.G1610A:p.R537H	exônica	não sinônimo	0/1		3,249E-05	VUS	0,473	sim
P062	F	curto	telomeropatia	TERT:NM_001193376:exon2:c.G1492A:p.G498R	exônica	não sinônimo	0/1			VUS	0,414	sim
P063	F	curto	telomeropatia	RTEL1:NM_001283009:exon34:c.3775_3776del:p.A1259Lfs*2	exônica	deleção com frameshift	0/1			LP	0,41	não
P064	M	curto	telomeropatia	TERC:NR_001566::n.314_315del	ncRNA		0/1			VUS	0,438	não
P065	M	curto	telomeropatia	TERC:NR_001566::n.36C>T	ncRNA		0/1			VUS	0,54	sim
P066	F	curto	telomeropatia	RTEL1:NM_001283010:exon33:c.2887_2890del:p.Q964Gfs*57	exônica	deleção com frameshift	0/1			LP	0,509	não
P067	M	curto	telomeropatia	TERC:NR_001566::n.94C>T	ncRNA		0/1			VUS	0,53	não
P067	M	curto	telomeropatia	TERT:NM_198253:exon11:c.C2701T:p.R901W	exônica	não sinônimo	0/1	rs199422304		VUS	0,5	sim
P068	M	curto	telomeropatia	RTEL1:NM_001283010:exon33:c.2887_2890del:p.Q964Gfs*57	exônica	deleção com frameshift	0/1			LP	0,435	não
P069	M	curto	telomeropatia	TINF2:NM_001099274:exon6:c.849dupC:p.T284Hfs*8	exônica	inserção com frameshift	0/1	rs199422315		LP	0,454	sim
P070	F	curto	telomeropatia	TERT:NM_001193376:exon3:c.T1639G:p.W547G	exônica	não sinônimo	0/1			VUS	0,462	sim
P070	F	curto	telomeropatia	RTEL1:NM_001283010:exon29:c.2209_2211del:p.H737del	exônica	deleção sem frameshift	0/1			VUS	0,432	não
P071	F	curto	telomeropatia	NOP10:NM_018648:exon2:c.C100T:p.R34W	exônica	não sinônimo	0/1	rs121908092		LP	0,469	sim
P071	F	curto	telomeropatia	RTEL1:NM_001283010:exon29:c.C2251T:p.R751X	exônica	ganho de parada	0/1	rs398123017	0,0001	P	0,487	sim
P072	M	curto	telomeropatia	RTEL1:NM_001283009:exon5:c.C438G:p.I146M	exônica	não sinônimo	0/1			VUS	0,548	não
P073	F	curto	telomeropatia	TERT:NM_198253:exon11:c.C2701T:p.R901W	exônica	não sinônimo	0/1	rs199422304		VUS	0,384	sim
P074	M	curto	telomeropatia	ZCCHC8:NM_001350935:exon6:c.602_604del:p.E201del	exônica	deleção sem frameshift	0/1	rs552902860	0,0086	VUS	0,382	sim
P075	M	curto	telomeropatia	TERT:NM_001193376:exon10:c.G2605A:p.D869N	exônica	não sinônimo	0/1			VUS	0,469	não

P076	F	curto	telomeropatia	PARN:NM_001242992:exon16:c.A964T:p.S322C	exônica	não sinônimo	0/1	rs138984302	0,0092	VUS	0,421	sim
P076	F	curto	telomeropatia	RTEL1:NM_001283010:exon23:c.G1382A:p.R461Q	exônica	não sinônimo	0/1	rs35640778	0,0185	VUS	0,455	sim
P077	F	curto	telomeropatia	ACD:NM_001082486:exon5:c.G706A:p.D236N	exônica	não sinônimo	0/1		2,978E-05	VUS	0,462	sim
P078	F	curto	telomeropatia	RTEL1:NM_001283010:exon11:c.G302A:p.R101H	exônica	não sinônimo	0/1	rs113684274	0,0011	VUS	0,5	sim
P079	M	curto	telomeropatia	TERC:NR_001566::n.110_113del	ncRNA		0/1	rs199422270		LP	0,489	sim
P080	M	curto	telomeropatia	TINF2:NM_001099274:exon6:c.G845A:p.R282H	exônica	não sinônimo	0/1	rs121918544		P	0,48	sim
P081	F	curto	telomeropatia	TINF2:NM_001099274:exon6:c.G845A:p.R282H	exônica	não sinônimo	0/1	rs121918544		P	0,485	sim
P082	M		ribosomopatia	RPL5:NM_000969:exon3:c.166_169del:p.N57Efs*12	exônica	deleção com frameshift	0/1			P	0,482	não
P083	M		ribosomopatia	RPS26:NM_001029:exon2:c.7_8insAG:p.K4Rfs*42	exônica	inserção com frameshift	0/1			LP	0,443	não
P084	F		ribosomopatia	RPS19:NM_001022:exon4:c.C280T:p.R94X	exônica	ganho de parada	0/1	rs61762293	0	P	0,498	sim
P085	F		ribosomopatia	RPL11:NM_000975::c.397-2A>G	splicing		0/1			LP	0,429	não
P086	F		ribosomopatia	RPS19:NM_001022:exon4:c.G185A:p.R62Q	exônica	não sinônimo	0/1			LP	0,482	sim
P087	F		ribosomopatia	RPS19:NM_001022:exon3:c.G167C:p.R56P	exônica	não sinônimo	0/1			VUS	0,457	não
P088	M		ribosomopatia	RPL15:NM_001253380:exon1:c.T56G:p.M19R	exônica	não sinônimo	0/1			VUS	0,47	não
P089	F		ribosomopatia	RPL5:NM_000969:exon3:c.166_169del:p.N57Efs*12	exônica	deleção com frameshift	0/1			P	0,485	não
P090	F		ribosomopatia	RPL11:NM_000975::c.507+2T>C	splicing		0/1			LP	0,483	não
P091	M		ribosomopatia	RPS26:NM_001029:exon2:c.6_9del:p.K4Efs*40	exônica	deleção com frameshift	0/1			P	0,531	não
P092	F		ribosomopatia	RPL15:NM_001253380:exon2:c.C274A:p.L92I	exônica	não sinônimo	0/1	rs146413350	0,001	VUS	0,511	sim
P093	M		ribosomopatia	RPS19:NM_001022:exon2:c.G3T:p.M1?	exônica	perda de início	0/1	rs138938035		P	0,429	sim
P094	F		ribosomopatia	RPS24:NM_001026:exon4:c.C308G:p.S103X	exônica	ganho de parada	0/1			LP	0,458	não
P095	F		ribosomopatia	RPS19:NM_001022:exon4:c.C184T:p.R62W	exônica	não sinônimo	0/1	rs104894711		LP	0,471	sim
P096	F		ribosomopatia	RPS26:NM_001029::c.312+1G>T	splicing		0/1			LP	0,44	não
P097	F		ribosomopatia	RPS26:NM_001029:exon2:c.6_9del:p.K4Efs*40	exônica	deleção com frameshift	0/1			P	0,44	não

P098	F	ribosomopatia	RPL5:NM_000969:exon5:c.514_517del;p.I173Lfs*38	exônica	deleção com frameshift	0/1		LP	0,495	não	
P099	M	ribosomopatia	RPS26:NM_001029:exon2:c.C124T;p.R42X	exônica	ganho de parada	0/1		LP	0,449	não	
P100	M	ribosomopatia	RPS19:NM_001022:exon4:c.173-10C>G;p.?	splicing		0/1		VUS	0,485	não	
P101	M	ribosomopatia	RPS26:NM_001029:exon1:c.3+1G>A;p.?	splicing		0 1	rs148622862	LP	0,483	sim	
P102	F	ribosomopatia	RPS26:NM_001029:exon2:c.175delT;p.F59Sfs*11	exônica	deleção com frameshift	0/1		LP	0,484	não	
P103	M	ribosomopatia	RPL5:NM_000969:exon6:c.C625T;p.R209C	exônica	não sinônimo	0/1	rs770095159	6,663E-05	VUS	0,463	sim
P104	F	ribosomopatia	RPS26:NM_001029::c.3+2T>C	splicing		0/1		LP	0,536	não	
P105	F	ribosomopatia	RPL5:NM_000969:exon3:c.172_173del;p.D59Yfs*53	exônica	deleção com frameshift	0/1		P	0,487	não	
P106	M	ribosomopatia	RPL35A:NM_000996:exon4:c.G266A;p.R89Q	exônica	não sinônimo	0/1	rs145244860	0,0083	VUS	0,466	sim
P107	F	ribosomopatia	RPS7:NM_001011:exon6:c.G455C;p.R152P	exônica	não sinônimo	0/1		VUS	0,493	sim	
P108	M	ribosomopatia	RPS19:NM_001022:exon4:c.324delG;p.L110*	exônica	ganho de parada	0/1		LP	0,379	não	
P109	M	AF	PALB2:NM_024675:exon12:c.C3236T;p.A1079V	exônica	não sinônimo	0/1	rs769342316	1,499E-05	VUS	0,446	sim
P109	M	AF	PALB2:NM_024675:exon8:c.T2816G;p.L939W	exônica	não sinônimo	0/1	rs45478192	0,0023	VUS	0,521	sim
P110	M	AF	FANCB:NM_152633:exon2:c.T52C;p.Y18H	exônica	não sinônimo	1/1		VUS	1	sim	
P111	M	AF	FANCA:NM_000135:exon32:c.G3164T;p.R1055L	exônica	não sinônimo	1/1		0,0001	LP	0,959	sim
P112	M	AF	FANCB:NM_152633:exon3:c.T989C;p.I330T	exônica	não sinônimo	1/1	rs200161949	0,002	VUS	0,995	sim
P113	M	AF	BRCA2:NM_000059:exon16:c.A7712G;p.E2571G	exônica	não sinônimo	0/1	rs55689095	0,0008	VUS	0,479	sim
P113	M	AF	BRCA2:NM_000059:exon22:c.A8830T;p.I2944F	exônica	não sinônimo	0/1	rs4987047	0,0419	VUS	0,455	sim
P114	M	AF	FANCG:NM_004629:exon9:c.1077-2A>G	splicing		1/1	rs769547477	8,944E-05	P	0,997	sim
P115	F	AF	FANCB:NM_152633:exon6:c.G1442A;p.R481H	exônica	não sinônimo	0/1	rs140198444	0,0017	VUS		sim
P115	F	AF	FANCB:NM_001018113::c.1105-2_1105-1insTATT	splicing		0/1	rs767651966	0,039	VUS		sim
P116	M	AF	FANCA:NM_000135:exon28:c.2604_2609del;p.Q869_F870del	exônica	deleção sem frameshift	1/1	rs745864224	1,502E-05	LP	0,98	sim
P117	F	AF	FANCE:NM_021922:exon2:c.519dupG;p.R176Qfs*12	exônica	inserção com frameshift	1/1	rs773363446	6,119E-05	LP	1	sim

P118	M	AF	SLX4:NM_032444:exon15:c.C5281T:p.R1761C	exônica	não sinônimo	0/1	rs143818824	0,031	VUS	0,51	sim
P118	M	AF	SLX4:NM_032444:exon12:c.G2359A:p.E787K	exônica	não sinônimo	0/1	rs140600202	0,0062	VUS	0,477	sim
P119	F	AF	BRCA2:NM_000059:exon11:c.G4183T:p.A1395S	exônica	não sinônimo	0/1	rs431825318		VUS	0,505	sim
P119	F	AF	BRCA2:NM_000059:exon22:c.G8851A:p.A2951T	exônica	não sinônimo	0/1	rs11571769	0,0418	VUS	0,437	sim
P120	F	AF	FANCA:NM_000135:exon38:c.3788_3790del:p.F1263del	exônica	deleção sem frameshift	1/1	rs397507553	0,0003	P	1	sim
P121	F	AF	RAD51C:NM_002876:exon2:c.G376A:p.A126T	exônica	não sinônimo	0/1	rs61758784	0,0099	VUS	0,447	sim
P121	F	AF	RAD51C:NM_058216:exon6:c.A859G:p.T287A	exônica	não sinônimo	0/1	rs28363317	0,016	VUS	0,495	sim
P122	M	NCG	G6PC3:NM_001319945:exon1:c.G136C:p.A46P	exônica	não sinônimo	1/1			VUS	1	não
P123	M	NCG	ELANE:NM_001972:exon5:c.618delG:p.V207Sfs*5	exônica	deleção com frameshift	0/1			LP	0,494	não
P124	F	NCG	ELANE:NM_001972:exon5:c.C684A:p.Y228X	exônica	ganho de parada	0/1			LP	0,466	não
P125	F	NCG	ELANE:NM_001972:exon4:c.C377T:p.S126L	exônica	não sinônimo	0/1	rs137854450		LP	0,462	sim
P126	M	NCG	GF11:NM_001127215::c.-100+2T>A	splicing		0/1			VUS	0,482	não
P127	M	síndrome de LIG4	LIG4:NM_001098268:exon2:c.1271_1275del:p.K424Rfs*20	exônica	deleção com frameshift	0/1	rs772226399	0,0003	P	0,552	sim
P127	M	síndrome de LIG4	LIG4:NM_001098268:exon2:c.T56G:p.L19W	exônica	não sinônimo	0/1	rs765520147	0,0000869	VUS	0,467	sim
P128	M	síndrome de LIG4	LIG4:NM_001098268:exon2:c.1751_1755del:p.I584Rfs*2	exônica	deleção com frameshift	0/1	rs770128720	9,619E-05	P	0,458	sim
P128	M	síndrome de LIG4	LIG4:NM_001098268:exon2:c.T56G:p.L19W	exônica	não sinônimo	0/1	rs765520147	0,0000869	VUS	0,377	sim
P129	M	síndrome de LIG4	LIG4:NM_001098268:exon2:c.T296A:p.L99X	exônica	ganho de parada	0/1			LP	0,412	não
P129	M	síndrome de LIG4	LIG4:NM_001098268:exon2:c.T56G:p.L19W	exônica	não sinônimo	0/1	rs765520147	0,0000869	VUS	0,455	sim
P130	F	deficiência de CTLA4	CTLA4:NM_001037631:exon2:c.C410T:p.P137L	exônica	não sinônimo	0/1			LP	0,417	sim
P131	M	deficiência de SAMD9	SAMD9:NM_001193307:exon2:c.T740C:p.I247T	exônica	não sinônimo	0/1	rs200540593	0,0002	VUS	0,469	sim

P132	M	deficiência de SAMD9	SAMD9:NM_001193307:exon2:c.T866A:p.F289Y	exônica	não sinônimo	0/1		VUS	0,431	não	
P133	M	deficiência de SAMD9	SAMD9:NM_001193307:exon2:c.C1337G:p.S446C	exônica	não sinônimo	0/1		VUS	0,438	sim	
P134	M	deficiência de SAMD9	SAMD9L:NM_001303500:exon4:c.3066delT:p.Y1022*	exônica	ganho de parada	0/1		LP	0,488	não	
P135	M	deficiência de GATA2	GATA2:NM_001145662:exon2:c.C216A:p.Y72X	exônica	ganho de parada	0/1		VUS	0,396	sim	
P136	M	deficiência de GATA2	GATA2:NM_001145662:exon5:c.G1040A:p.R347H	exônica	não sinônimo	0/1		LP	0,456	sim	
P137	F	deficiência de GATA2	GATA2:NM_001145662:exon3:c.C379A:p.H127N	exônica	não sinônimo	0/1		VUS	0,515	sim	
P138	M	deficiência de GATA2	GATA2:NM_001145662:exon3:c.C610T:p.R204X	exônica	ganho de parada	0/1		P	0,49	sim	
P139	F	deficiência de GATA2	GATA2:NM_001145662:exon3:c.A661C:p.M221L	exônica	não sinônimo	0/1	8,962E-06	VUS	0,456	sim	
P140	M	deficiência de GATA2	GATA2:NM_001145662:exon6:c.C1144T:p.R382W	exônica	não sinônimo	0/1		P	0,518	sim	
P141	F	deficiência de MECOM	MECOM:NM_001105078:exon8:c.1954delC:p.E653Kfs*3	exônica	deleção com frameshift	0/1		LP	0,466	não	
P142	F	deficiência de MECOM	MECOM:NM_001105078:exon7:c.G1090A:p.G364R	exônica	não sinônimo	0/1	rs746379599	0,0003	VUS	0,481	sim
P143	M	deficiência de MECOM	MECOM:NM_001105078:exon7:c.T1907G:p.F636C	exônica	não sinônimo	0/1		0,0002	VUS	0,507	sim
P143	M	deficiência de MECOM	MECOM:NM_001105078:exon7:c.C737T:p.A246V	exônica	não sinônimo	0/1		0,0001	VUS	0,496	sim
P144	M	TAC	MPL:NM_005373:exon3:c.G305C:p.R102P	exônica	não sinônimo	1/1	rs28928907	0,0082	P	0,985	sim
P145	M	TAC	MPL:NM_005373:exon3:c.G305C:p.R102P	exônica	não sinônimo	0/1	rs28928907	0,0082	P	0,49	sim
P145	M	TAC	MPL:NM_005373:exon8:c.G1305C:p.W435C	exônica	não sinônimo	0/1		0,0002	LP	0,545	sim
P146	M	TAC	MPL:NM_005373:exon3:c.G305C:p.R102P	exônica	não sinônimo	1/1	rs28928907	0,0082	P	0,967	sim
P147	F	curto	inconclusivo								
P148	F	curto	inconclusivo								
P149	F	normal	inconclusivo								
P150	F	normal	inconclusivo								

P151	M	curto	inconclusivo
P152	F	curto	inconclusivo
P153	F	normal	inconclusivo
P154	M	normal	inconclusivo
P155	F	normal	inconclusivo
P156	M	curto	inconclusivo
P157	F	normal	inconclusivo
P158	F	normal	inconclusivo
P159	F	curto	inconclusivo
P160	M	normal	inconclusivo
P161	M	curto	inconclusivo
P162	F	curto	inconclusivo
P163	F	normal	inconclusivo
P164	F	normal	inconclusivo
P165	F	normal	inconclusivo
P166	M	normal	inconclusivo
P167	F	curto	inconclusivo
P168	M	curto	inconclusivo
P169	M	normal	inconclusivo
P170	F	curto	inconclusivo
P171	M	normal	inconclusivo
P172	F	normal	inconclusivo
P173	M	curto	inconclusivo
P174	M	normal	inconclusivo
P175	F	curto	inconclusivo
P176	F	curto	inconclusivo
P177	F	normal	inconclusivo
P178	M	curto	inconclusivo
P179	F	normal	inconclusivo
P180	M	curto	inconclusivo
P181	F	curto	inconclusivo
P182	M	curto	inconclusivo

P183	F	curto	inconclusivo
P184	M		inconclusivo
P185	F	normal	inconclusivo
P186	F	curto	inconclusivo
P187	F	curto	inconclusivo
P188	F	curto	inconclusivo
P189	F	curto	inconclusivo
P190	M		inconclusivo
P191	M	curto	inconclusivo
P192	F	normal	inconclusivo
P193	M	normal	inconclusivo
P194	F	curto	inconclusivo
P195	M	curto	inconclusivo
P196	M	curto	inconclusivo
P197	F	curto	inconclusivo
P198	F	curto	inconclusivo
P199	F	normal	inconclusivo
P200	F	curto	inconclusivo
P201	M	curto	inconclusivo
P202	M	normal	inconclusivo
P203	M	normal	inconclusivo
P204	M	curto	inconclusivo
P205	F	curto	inconclusivo
P206	M	curto	inconclusivo
P207	M	curto	inconclusivo
P208	F	normal	inconclusivo
P209	M	curto	inconclusivo
P210	M		inconclusivo
P211	M	normal	inconclusivo
P212	F	normal	inconclusivo
P213	F	curto	inconclusivo
P214	F	normal	inconclusivo

P215	F	normal	inconclusivo
P216	F	normal	inconclusivo
P217	M	curto	inconclusivo
P218	M	normal	inconclusivo
P219	F	curto	inconclusivo
P220	F	curto	inconclusivo
P221	M	normal	inconclusivo
P222	M	normal	inconclusivo
P223	M	normal	inconclusivo
P224	M	curto	inconclusivo
P225	M	normal	inconclusivo
P226	F	normal	inconclusivo
P227	M	normal	inconclusivo
P228	F	normal	inconclusivo
P229	M	normal	inconclusivo
P230	M	normal	inconclusivo
P231	M	curto	inconclusivo
P232	M	curto	inconclusivo
P233	M	normal	inconclusivo
P234	M	curto	inconclusivo
P235	F	curto	inconclusivo
P236	F	normal	inconclusivo
P237	F	curto	inconclusivo
P238	M	curto	inconclusivo
P239	F	normal	inconclusivo
P240	M	normal	inconclusivo
P241	F	normal	inconclusivo
P242	F	curto	inconclusivo
P243	F	curto	inconclusivo
P244	F	normal	inconclusivo
P245	M	normal	inconclusivo
P246	F	normal	inconclusivo

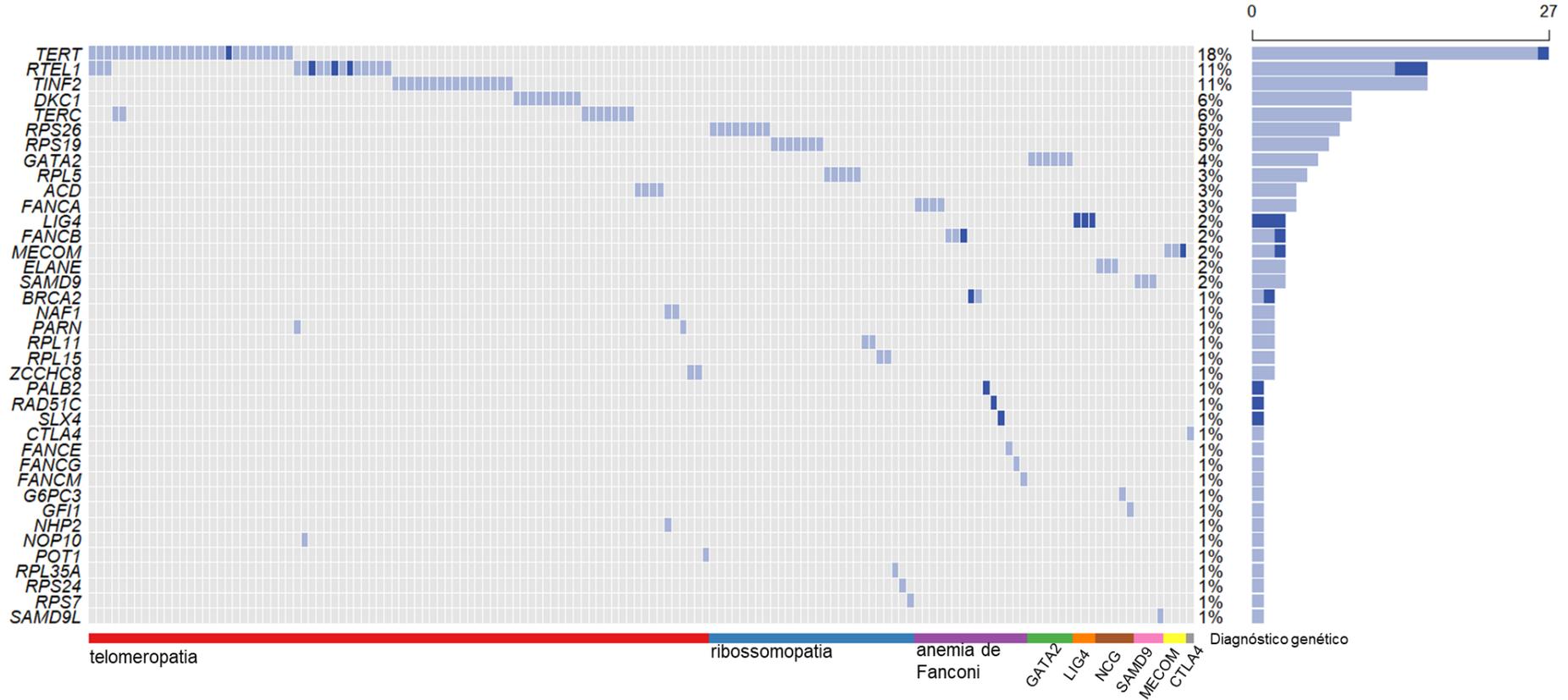
P247	F	curto	inconclusivo
P248	F	curto	inconclusivo
P249	F	curto	inconclusivo
P250	F	curto	inconclusivo
P251	M	normal	inconclusivo
P252	F		inconclusivo
P253	M	curto	inconclusivo
P254	M	normal	inconclusivo
P255	F	curto	inconclusivo
P256	F	curto	inconclusivo
P257	F	curto	inconclusivo
P258	M		inconclusivo
P259	F	curto	inconclusivo
P260	M	normal	inconclusivo
P261	F	curto	inconclusivo
P262	F	curto	inconclusivo
P263	M	normal	inconclusivo
P264	M	normal	inconclusivo
P265	F	normal	inconclusivo
P266	F		inconclusivo
P267	F		inconclusivo
P268	F		inconclusivo
P269	M	curto	inconclusivo
P270	M	normal	inconclusivo
P271	M	normal	inconclusivo
P272	F	curto	inconclusivo
P273	M	curto	inconclusivo
P274	M		inconclusivo
P275	F	normal	inconclusivo
P276	M	normal	inconclusivo
P277	F	curto	inconclusivo
P278	M	normal	inconclusivo

P279	F	curto	inconclusivo
P280	M	normal	inconclusivo
P281	F	normal	inconclusivo
P282	F	normal	inconclusivo
P283	M		inconclusivo
P284	M	curto	inconclusivo
P285	F	curto	inconclusivo
P286	M	normal	inconclusivo
P287	M	normal	inconclusivo
P288	M	normal	inconclusivo
P289	M	normal	inconclusivo
P290	F		inconclusivo
P291	F	curto	inconclusivo
P292	F	curto	inconclusivo
P293	M	curto	inconclusivo
P294	M	curto	inconclusivo
P295	F	curto	inconclusivo
P296	M	curto	inconclusivo
P297	F	curto	inconclusivo
P298	M	normal	inconclusivo
P299	F	normal	inconclusivo
P300	F	normal	inconclusivo
P301	F	curto	inconclusivo
P302	F	normal	inconclusivo
P303	F	curto	inconclusivo
P304	F	normal	inconclusivo
P305	M		inconclusivo
P306	M	curto	inconclusivo
P307	M	curto	inconclusivo
P308	M	curto	inconclusivo
P309	M	normal	inconclusivo
P310	F	normal	inconclusivo

P311	F	normal	inconclusivo
P312	M	curto	inconclusivo
P313	M	normal	inconclusivo
P314	M	normal	inconclusivo

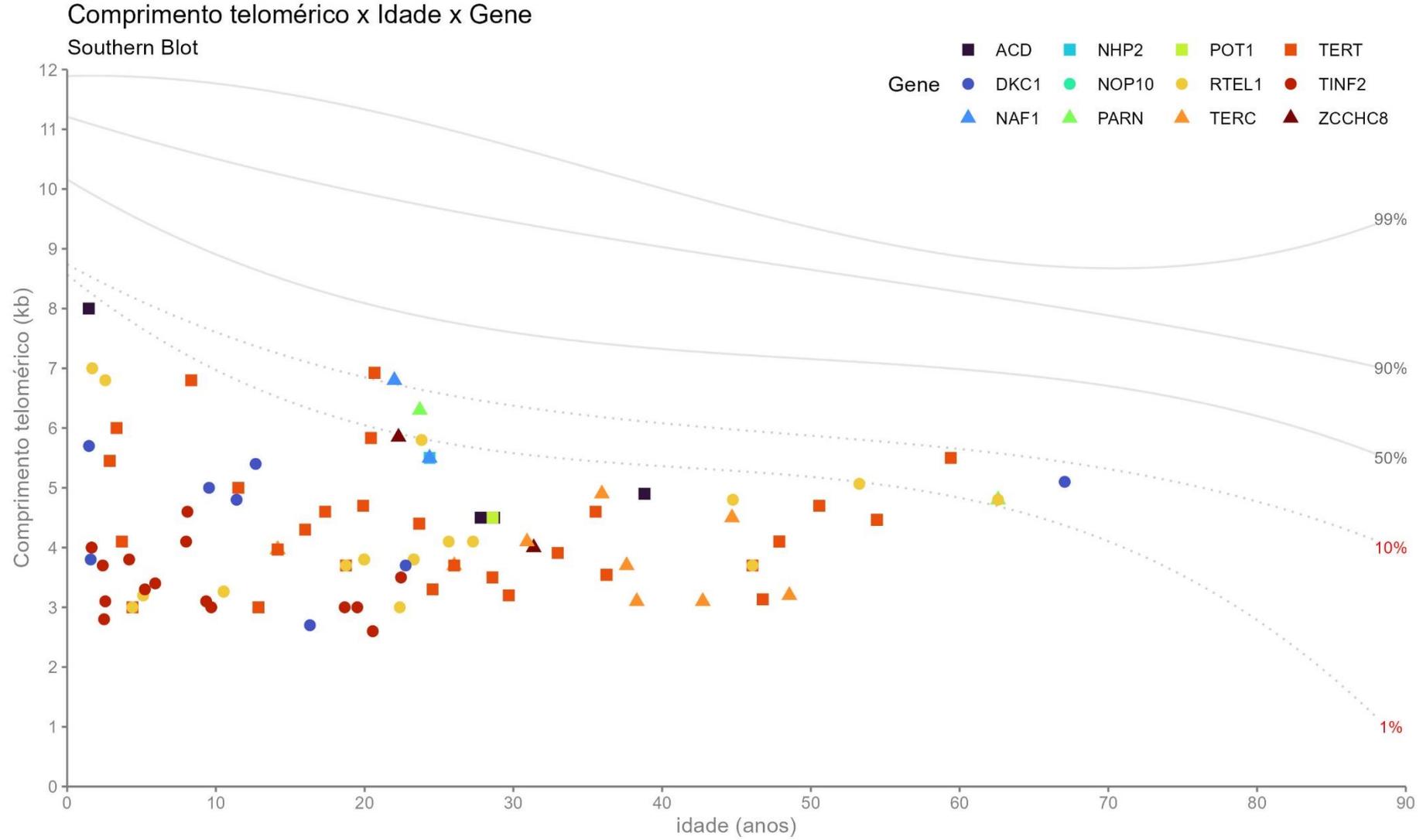
Abreviações: M (masculino), F (feminino), FeqMaxPop (frequência máxima da variante em diferentes populações).

APÊNDICE C - Genes alterados em pacientes (N = 146) com diagnóstico genético de síndrome de falência medular hereditária



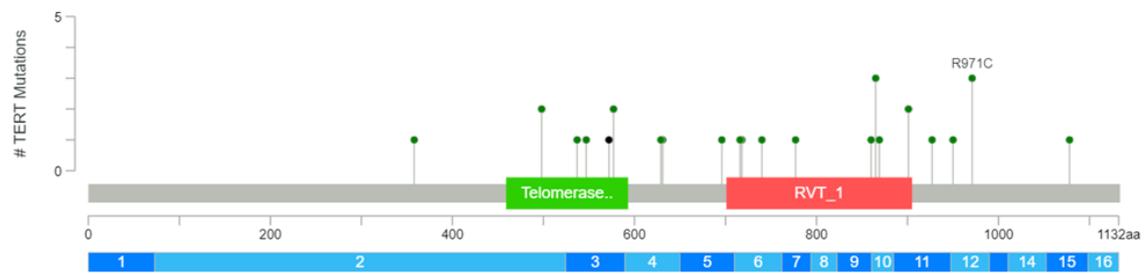
Abreviações: NCG (neutropenia congênita grave), SAMD9 (deficiência de SAMD9), MECOM (deficiência de MECOM), CTLA4 (deficiência de CTLA4), GATA2 (deficiência de GATA2) e LIG4 (deficiência de LIG4)

APÊNDICE D – Comprimento teloméricos de paciente com diagnóstico genético de telomeropatia

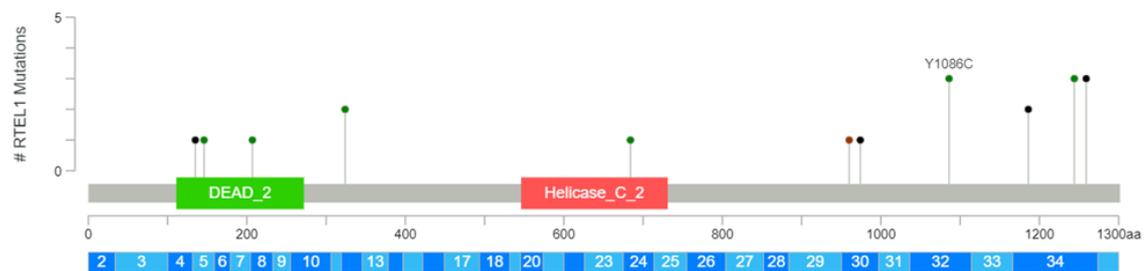


APÊNDICE E – Gráficos de *lollipop* dos dez genes com mais variantes

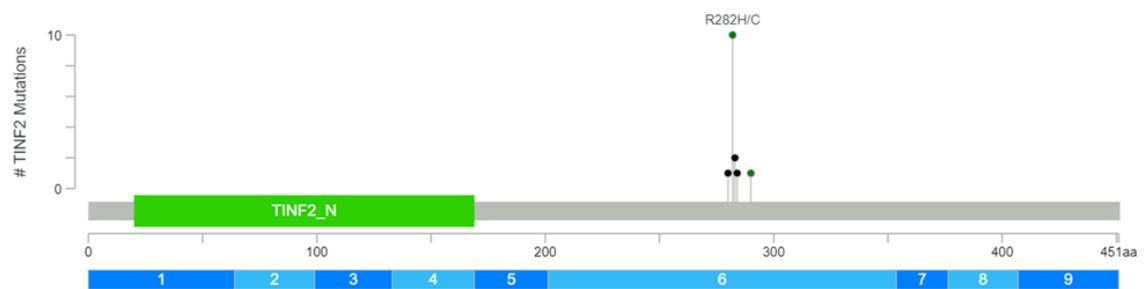
TERT



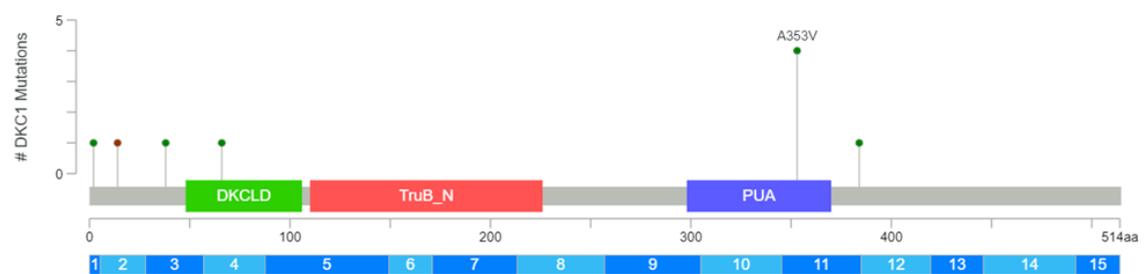
RTEL1



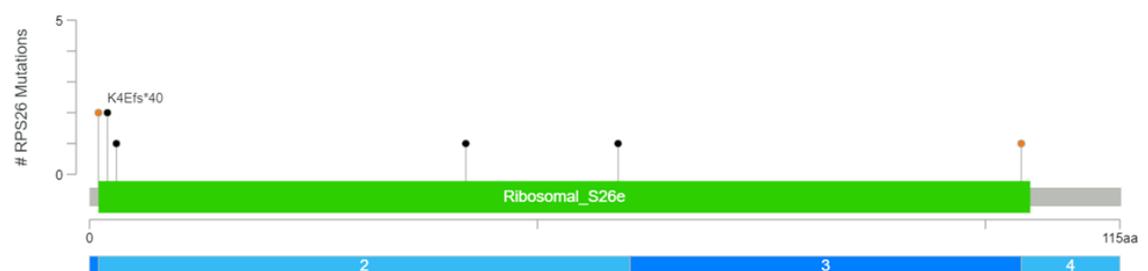
TINF2

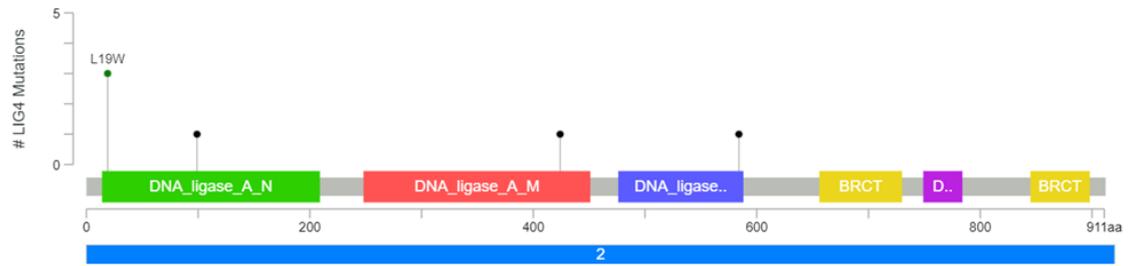
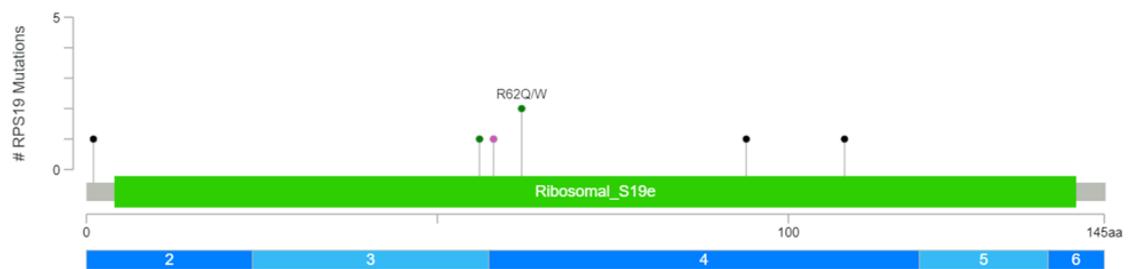
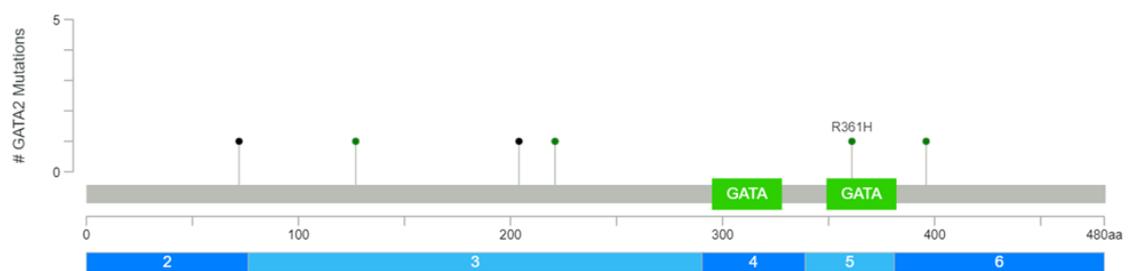
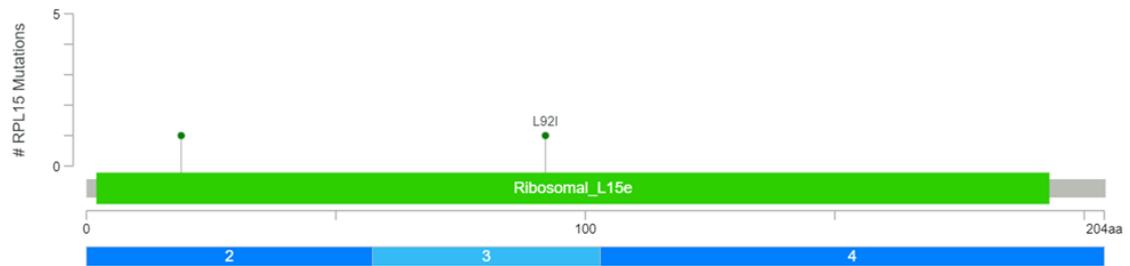
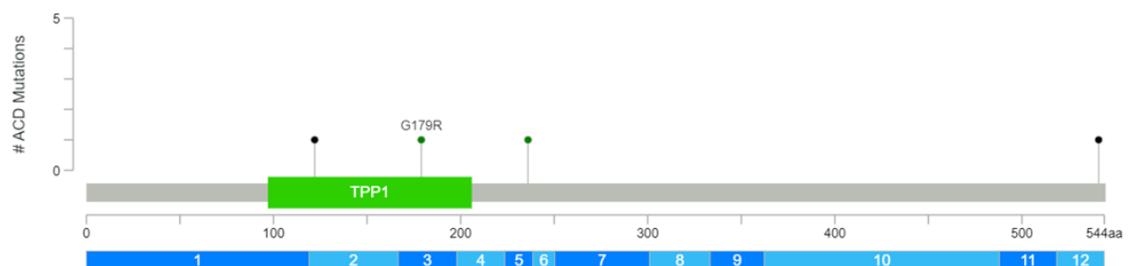


DKC1



RPS26



LIG4**RPS19****GATA2****RPL15****ACD**

APÊNDICE F – Código em SQL do modelo físico

```
-- MySQL Script generated by MySQL Workbench
-- Sat Nov 11 19:52:45 2023
-- Model: New Model    Version: 1.0
-- MySQL Workbench Forward Engineering

SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@@SQL_MODE, SQL_MODE='ONLY_FULL_GROUP_BY,STRICT_TRANS_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_ZERO,NO_ENGINE_SUBSTITUTION';

-----
-- Schema bmf_db
-----

-----
-- Schema bmf_db
-----
CREATE SCHEMA IF NOT EXISTS `bmf_db` DEFAULT CHARACTER SET utf8 ;
-----
-- Schema bmf_db
-----

-----
-- Schema bmf_db
-----
CREATE SCHEMA IF NOT EXISTS `bmf_db` DEFAULT CHARACTER SET utf8 COLLATE
utf8_unicode_ci ;
USE `bmf_db` ;

-----
-- Table `mydb`.`timestamps`
-----
CREATE TABLE IF NOT EXISTS `bmf_db`.`timestamps` (
  `create_time` TIMESTAMP NULL DEFAULT CURRENT_TIMESTAMP,
  `update_time` TIMESTAMP NULL);

USE `bmf_db` ;

-----
-- Table `bmf_db`.`clinicalEvolution`
-----
CREATE TABLE IF NOT EXISTS `bmf_db`.`clinicalEvolution` (
  `IdfollowUp` INT NOT NULL AUTO_INCREMENT,
  `DatefollowUp` DATE NULL,
  `IdMd` VARCHAR(45) NULL DEFAULT NULL,
  `SymptomSign1` INT NULL DEFAULT NULL,
  `SymptomSign2` VARCHAR(45) CHARACTER SET 'utf8' COLLATE 'utf8_unicode_ci'
  NULL DEFAULT NULL,
  `SymptomSign3` VARCHAR(45) CHARACTER SET 'utf8' COLLATE 'utf8_unicode_ci'
  NULL DEFAULT NULL,
  `SymptomSign4` VARCHAR(45) CHARACTER SET 'utf8' COLLATE 'utf8_unicode_ci'
  NULL DEFAULT NULL,
  `SymptomSign5` VARCHAR(45) CHARACTER SET 'utf8' COLLATE 'utf8_unicode_ci'
  NULL DEFAULT NULL,
  `DxHypothesis1` INT NULL DEFAULT NULL,
  `DxHypothesis2` INT NULL DEFAULT NULL,
  `DxHypothesis3` INT NULL DEFAULT NULL,
  `FinalDx` INT NULL DEFAULT NULL,
```

```

`Cid` VARCHAR(45) NULL DEFAULT NULL,
`DeathDate` DATE NULL DEFAULT NULL,
`Tx` VARCHAR(45) NULL,
`DateTx` VARCHAR(45) NULL,
`ResponseTx` VARCHAR(45) NULL,
`Bmt` VARCHAR(45) NULL,
`DateBmt` VARCHAR(45) NULL,
`Death` VARCHAR(45) NULL,
`DateDeath` VARCHAR(45) NULL,
`CauseDeath` VARCHAR(45) NULL,
PRIMARY KEY (`IdfollowUp`))
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_bin;

```

```

-----
-- Table `bmf_db`.`patientInfo`
-----

```

```

CREATE TABLE IF NOT EXISTS `bmf_db`.`patientInfo` (
  `IdPatient` VARCHAR(10) NOT NULL,
  `IdPatientOther` VARCHAR(15) CHARACTER SET 'utf8' NULL DEFAULT NULL,
  `NamePatient` VARCHAR(45) NULL,
  `NumberPatient` VARCHAR(45) NULL,
  `Sex` VARCHAR(15) CHARACTER SET 'utf8' NULL DEFAULT NULL,
  `DobPatient` DATE NULL DEFAULT NULL,
  `CityBornPatient` VARCHAR(15) CHARACTER SET 'utf8' NULL DEFAULT NULL,
  `StateBornPatient` VARCHAR(15) CHARACTER SET 'utf8' NULL DEFAULT NULL,
  `IdHosp` VARCHAR(15) CHARACTER SET 'utf8' NULL DEFAULT NULL,
  `IdMd` VARCHAR(45) NULL,
  `clinicalEvolution_IdfollowUp` INT NOT NULL,
  PRIMARY KEY (`IdPatient`, `clinicalEvolution_IdfollowUp`),
  INDEX
    (`clinicalEvolution_IdfollowUp` ASC) VISIBLE,
  CONSTRAINT `fk_patientInfo_clinicalEvolution1`
    FOREIGN KEY (`clinicalEvolution_IdfollowUp`)
    REFERENCES `bmf_db`.`clinicalEvolution` (`IdfollowUp`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_bin;

```

```

-----
-- Table `bmf_db`.`sample`
-----

```

```

CREATE TABLE IF NOT EXISTS `bmf_db`.`sample` (
  `IdSample` VARCHAR(20) NOT NULL,
  `IdPatient` VARCHAR(10) NULL,
  `IdSampleOther` VARCHAR(45) NULL,
  `AgeSample` VARCHAR(45) NULL,
  `DateSample` VARCHAR(45) NULL DEFAULT NULL,
  `SourceSample` VARCHAR(45) NULL DEFAULT NULL,
  `IdDna` VARCHAR(45) NULL,
  `IdRna` VARCHAR(45) NULL,
  `IdCell` VARCHAR(45) NULL,
  `IdFib` VARCHAR(45) NULL,
  `idPlasm` VARCHAR(45) NULL,
  `study_has_patientInfo_study_IdStudy` INT NULL,
  `study_has_patientInfo_patientInfo_IdPatient` INT NULL,

```

```

`patientInfo_IdPatient` VARCHAR(10) NOT NULL,
PRIMARY KEY (`IdSample`, `patientInfo_IdPatient`),
INDEX `fk_sample_patientInfo1_idx` (`patientInfo_IdPatient` ASC) VISIBLE)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;

```

```

-----
-- Table `bmf_db`.`cellSample`
-----

```

```

CREATE TABLE IF NOT EXISTS `bmf_db`.`cellSample` (
  `IdCell` INT NOT NULL,
  `IdSample` VARCHAR(20) NOT NULL,
  `DataProcCell` DATE NULL DEFAULT NULL,
  `TypeCell` VARCHAR(45) NULL DEFAULT NULL,
  `QuantCell` INT NULL,
  `StorageCell` VARCHAR(45) NULL DEFAULT NULL,
  `sample_IdSample` VARCHAR(20) NOT NULL,
  `sample_patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  PRIMARY KEY (`IdCell`, `sample_IdSample`, `sample_patientInfo_IdPatient`,
`IdSample`),
  INDEX `fk_cellSample_sample1_idx` (`sample_IdSample` ASC,
`sample_patientInfo_IdPatient` ASC) VISIBLE,
  CONSTRAINT `fk_cellSample_sample1`
    FOREIGN KEY (`sample_IdSample`, `sample_patientInfo_IdPatient`)
    REFERENCES `bmf_db`.`sample` (`IdSample`, `patientInfo_IdPatient`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;

```

```

-----
-- Table `bmf_db`.`dnaSample`
-----

```

```

CREATE TABLE IF NOT EXISTS `bmf_db`.`dnaSample` (
  `IdDna` VARCHAR(20) NOT NULL,
  `IdSample` VARCHAR(20) NOT NULL,
  `DataProcDna` DATE NULL DEFAULT NULL,
  `KitExtraction` VARCHAR(45) NULL DEFAULT NULL,
  `QuantNano` VARCHAR(45) NULL DEFAULT NULL,
  `Ratio260_230` VARCHAR(45) NULL DEFAULT NULL,
  `Ratio260_280` VARCHAR(45) NULL DEFAULT NULL,
  `QuantFluor` VARCHAR(45) NULL DEFAULT NULL,
  `Integrity` VARCHAR(45) NULL DEFAULT NULL,
  `StorageDna` VARCHAR(45) NULL DEFAULT NULL,
  `IdPatient` INT NULL,
  `patientInfo_IdPatient` INT NULL,
  `sample_IdSample` VARCHAR(20) NOT NULL,
  `sample_patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  PRIMARY KEY (`IdDna`, `sample_IdSample`, `sample_patientInfo_IdPatient`,
`IdSample`),
  INDEX `fk_dnaSample_sample1_idx` (`sample_IdSample` ASC,
`sample_patientInfo_IdPatient` ASC) VISIBLE,
  CONSTRAINT `fk_dnaSample_sample1`
    FOREIGN KEY (`sample_IdSample`, `sample_patientInfo_IdPatient`)
    REFERENCES `bmf_db`.`sample` (`IdSample`, `patientInfo_IdPatient`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)

```

```
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;
```

```
-----
-- Table `bmf_db`.`fibSample`
-----
```

```
CREATE TABLE IF NOT EXISTS `bmf_db`.`fibSample` (
  `IdFib` INT NOT NULL,
  `IdSample` VARCHAR(20) NULL DEFAULT NULL,
  `DataProcFibro` DATE NULL DEFAULT NULL,
  `QuantFibro` INT NULL DEFAULT NULL,
  `StorageFibro` VARCHAR(45) NULL DEFAULT NULL,
  `IdDna` VARCHAR(45) NULL DEFAULT NULL,
  `IdRna` VARCHAR(45) NULL,
  `sample_IdSample` VARCHAR(20) NOT NULL,
  `sample_patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  PRIMARY KEY (`IdFib`, `sample_IdSample`, `sample_patientInfo_IdPatient`),
  UNIQUE INDEX `IdFib_UNIQUE` (`IdFib` ASC) VISIBLE,
  INDEX `fk_fibSample_sample1_idx` (`sample_IdSample` ASC,
  `sample_patientInfo_IdPatient` ASC) VISIBLE,
  CONSTRAINT `fk_fibSample_sample1`
    FOREIGN KEY (`sample_IdSample`, `sample_patientInfo_IdPatient`)
    REFERENCES `bmf_db`.`sample` (`IdSample`, `patientInfo_IdPatient`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;
```

```
-----
-- Table `bmf_db`.`hospitalInfo`
-----
```

```
CREATE TABLE IF NOT EXISTS `bmf_db`.`hospitalInfo` (
  `IdHosp` VARCHAR(10) NOT NULL,
  `NameHosp` VARCHAR(45) NULL DEFAULT NULL,
  `CityHosp` VARCHAR(45) NULL DEFAULT NULL,
  `StateHosp` VARCHAR(45) NULL DEFAULT NULL,
  `AddressHosp` VARCHAR(45) NULL DEFAULT NULL,
  PRIMARY KEY (`IdHosp`))
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;
```

```
-----
-- Table `bmf_db`.`physicianInfo`
-----
```

```
CREATE TABLE IF NOT EXISTS `bmf_db`.`physicianInfo` (
  `IdMd` VARCHAR(10) NOT NULL,
  `NumberMd` VARCHAR(45) NULL,
  `NameMd` VARCHAR(45) NULL DEFAULT NULL,
  `EmailMd` VARCHAR(45) NULL DEFAULT NULL,
  `TelMd` VARCHAR(45) NULL DEFAULT NULL,
  `IdHosp` INT NULL DEFAULT NULL,
  PRIMARY KEY (`IdMd`))
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8mb3;
```

```
-----
-- Table `bmf_db`.`plasmSample`
-----
```

```
CREATE TABLE IF NOT EXISTS `bmf_db`.`plasmSample` (
  `IdPlasm` INT NOT NULL,
  `IdSample` VARCHAR(20) NULL DEFAULT NULL,
  `DataProcPlasm` VARCHAR(45) NULL DEFAULT NULL,
  `StoragePlasm` VARCHAR(45) NULL DEFAULT NULL,
  `ExtractSample` VARCHAR(45) NULL DEFAULT NULL,
  `sample_IdSample` INT NOT NULL,
  `sample_patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  PRIMARY KEY (`IdPlasm`, `sample_IdSample`,
`sample_patientInfo_IdPatient`),
  INDEX `fk_plasmSample_sample1_idx` (`sample_IdSample` ASC,
`sample_patientInfo_IdPatient` ASC) VISIBLE)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;
```

```
-----
-- Table `bmf_db`.`study`
-----
```

```
CREATE TABLE IF NOT EXISTS `bmf_db`.`study` (
  `IdStudy` INT UNSIGNED NOT NULL,
  `Title` VARCHAR(100) NULL,
  `RespResearcher1` VARCHAR(45) NULL DEFAULT NULL,
  `RespResearcher2` VARCHAR(45) NULL DEFAULT NULL,
  `RespResearcher3` VARCHAR(45) NULL DEFAULT NULL,
  `EcaNumber` VARCHAR(45) NULL DEFAULT NULL,
  `IdPatient` VARCHAR(10) NULL,
  `patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  PRIMARY KEY (`IdStudy`, `patientInfo_IdPatient`),
  INDEX `fk_study_patientInfo1_idx` (`patientInfo_IdPatient` ASC) VISIBLE)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;
```

```
-----
-- Table `bmf_db`.`flowFish`
-----
```

```
CREATE TABLE IF NOT EXISTS `bmf_db`.`flowFish` (
  `ffId` INT NOT NULL,
  `IdSample` VARCHAR(20) NULL DEFAULT NULL,
  `IdPatient` VARCHAR(10) NULL,
  `DateTelFf` DATE NULL DEFAULT NULL,
  `TrfExpectedFf` DECIMAL(2,2) NULL DEFAULT NULL,
  `TrfObserved` DECIMAL(2,2) NULL DEFAULT NULL,
  `SampleAgeFf` DECIMAL(2,2) NULL DEFAULT NULL,
  `ValorAgeFf` DECIMAL(2,2) NULL DEFAULT NULL,
  `DpNumberFf` DECIMAL(2,2) NULL DEFAULT NULL,
  `TrfPercentile` VARCHAR(10) NULL DEFAULT NULL,
  `method` VARCHAR(45) NULL,
  `cellSample_IdCell` INT NOT NULL,
  `cellSample_sample_IdSample` VARCHAR(20) NOT NULL,
  `cellSample_sample_patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  `cellSample_IdSample` VARCHAR(20) NOT NULL,
  PRIMARY KEY (`ffId`, `cellSample_IdCell`, `cellSample_sample_IdSample`,
`cellSample_sample_patientInfo_IdPatient`, `cellSample_IdSample`),
```

```

INDEX      `fk_flowFish_cellSample1_idx`      (`cellSample_IdCell`      ASC,
`cellSample_sample_IdSample` ASC, `cellSample_sample_patientInfo_IdPatient`
ASC, `cellSample_IdSample` ASC) VISIBLE,
CONSTRAINT `fk_flowFish_cellSample1`
FOREIGN KEY (`cellSample_IdCell` , `cellSample_sample_IdSample` ,
`cellSample_sample_patientInfo_IdPatient` , `cellSample_IdSample`)
REFERENCES `bmf_db`.`cellSample` (`IdCell` , `sample_IdSample` ,
`sample_patientInfo_IdPatient` , `IdSample`)
ON DELETE NO ACTION
ON UPDATE NO ACTION)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;

```

```

-----
-- Table `bmf_db`.`rnaSample`
-----

```

```

CREATE TABLE IF NOT EXISTS `bmf_db`.`rnaSample` (
  `IdRna` INT NOT NULL AUTO_INCREMENT,
  `IdSample` INT NULL DEFAULT NULL,
  `DataProcRna` DATE NULL DEFAULT NULL,
  `KitExtraction` VARCHAR(45) NULL DEFAULT NULL,
  `QuantNano` VARCHAR(45) NULL DEFAULT NULL,
  `Ratio260_230` VARCHAR(45) NULL DEFAULT NULL,
  `Ratio260_280` VARCHAR(45) NULL DEFAULT NULL,
  `QuantFluor` VARCHAR(45) NULL DEFAULT NULL,
  `Integrity` VARCHAR(45) NULL DEFAULT NULL,
  `StorageRna` VARCHAR(45) NULL DEFAULT NULL,
  PRIMARY KEY (`IdRna`),
  UNIQUE INDEX `IdRna_UNIQUE` (`IdRna` ASC) VISIBLE)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;

```

```

-----
-- Table `bmf_db`.`shouthernBlot`
-----

```

```

CREATE TABLE IF NOT EXISTS `bmf_db`.`shouthernBlot` (
  `sbId` INT NOT NULL,
  `IdPatient` VARCHAR(10) NULL,
  `IdSample` VARCHAR(20) NULL DEFAULT NULL,
  `DateTel` DATE NULL DEFAULT NULL,
  `TrfExpect` DECIMAL(2,2) NULL DEFAULT NULL,
  `TrfObserved` DECIMAL(2,2) NULL DEFAULT NULL,
  `SampleAge` DECIMAL(2,2) NULL DEFAULT NULL,
  `ValorAge` DECIMAL(2,2) NULL DEFAULT NULL,
  `DpNumber` DECIMAL(2,2) NULL DEFAULT NULL,
  `TrfPercentile` VARCHAR(10) NULL DEFAULT NULL,
  `method` VARCHAR(45) NULL,
  `dnaSample_IdDna` VARCHAR(20) NOT NULL,
  `dnaSample_sample_IdSample` VARCHAR(20) NOT NULL,
  `dnaSample_sample_patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  PRIMARY KEY (`sbId`, `dnaSample_IdDna`, `dnaSample_sample_IdSample`,
`dnaSample_sample_patientInfo_IdPatient`),
  INDEX      `fk_shouthernBlot_dnaSample1_idx`      (`dnaSample_IdDna`      ASC,
`dnaSample_sample_IdSample` ASC, `dnaSample_sample_patientInfo_IdPatient`
ASC) VISIBLE,
CONSTRAINT `fk_shouthernBlot_dnaSample1`

```

```

        FOREIGN KEY (`dnaSample_IdDna` , `dnaSample_sample_IdSample` ,
`dnaSample_sample_patientInfo_IdPatient`)
        REFERENCES `bmf_db`.`dnaSample` (`IdDna` , `sample_IdSample` ,
`sample_patientInfo_IdPatient`)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;

-----
-- Table `bmf_db`.`telSb`
-----
CREATE TABLE IF NOT EXISTS `bmf_db`.`telSb` (
  `IdTelFf` INT NOT NULL,
  `IdSample` INT NULL DEFAULT NULL,
  `DateTelSb` DATE NULL DEFAULT NULL,
  `TrfExpected` DECIMAL(2,2) NULL DEFAULT NULL,
  `TrfObserved` DECIMAL(2,2) NULL DEFAULT NULL,
  `AgeSampleSb` DECIMAL(2,2) NULL DEFAULT NULL,
  `ValorAgeSb` DECIMAL(2,2) NULL DEFAULT NULL,
  `DpNumbeSb` DECIMAL(2,2) NULL DEFAULT NULL,
  `TrfPercentileSb` VARCHAR(10) NULL DEFAULT NULL,
  `dnaSample_IdDna` INT NOT NULL,
  PRIMARY KEY (`IdTelFf`),
  INDEX `fk_telSb_dnaSample1_idx` (`dnaSample_IdDna` ASC) VISIBLE)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;

-----
-- Table `bmf_db`.`patientInfo_has_physicianInfo`
-----
CREATE TABLE IF NOT EXISTS `bmf_db`.`patientInfo_has_physicianInfo` (
  `patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  `physicianInfo_IdMd` VARCHAR(10) NOT NULL,
  PRIMARY KEY (`patientInfo_IdPatient`, `physicianInfo_IdMd`),
  INDEX `fk_patientInfo_has_physicianInfo_physicianInfo1_idx`
(`physicianInfo_IdMd` ASC) VISIBLE,
  INDEX `fk_patientInfo_has_physicianInfo_patientInfo1_idx`
(`patientInfo_IdPatient` ASC) VISIBLE)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_bin;

-----
-- Table `bmf_db`.`hospitalInfo_has_physicianInfo`
-----
CREATE TABLE IF NOT EXISTS `bmf_db`.`hospitalInfo_has_physicianInfo` (
  `hospitalInfo_IdHosp` VARCHAR(10) NOT NULL,
  `physicianInfo_IdMd` VARCHAR(10) NOT NULL,
  PRIMARY KEY (`hospitalInfo_IdHosp`, `physicianInfo_IdMd`),
  INDEX `fk_hospitalInfo_has_physicianInfo_physicianInfo1_idx`
(`physicianInfo_IdMd` ASC) VISIBLE,
  INDEX `fk_hospitalInfo_has_physicianInfo_hospitalInfo1_idx`
(`hospitalInfo_IdHosp` ASC) VISIBLE,
  UNIQUE INDEX `hospitalInfo_IdHosp_UNIQUE` (`hospitalInfo_IdHosp` ASC)
VISIBLE)

```

```
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;
```

```
-----
-- Table `bmf_db`.`variantFiltred`
-----
CREATE TABLE IF NOT EXISTS `bmf_db`.`variantFiltred` (
  `IdPatient` VARCHAR(10) NOT NULL,
  `IdSample` VARCHAR(20) NOT NULL,
  `Panel` INT NOT NULL,
  `Gene.refGene` VARCHAR(45) NULL DEFAULT NULL,
  `ExonicFunc.refGene` VARCHAR(20) NULL DEFAULT NULL,
  `AAChange.refGene` VARCHAR(200) NULL DEFAULT NULL,
  `etiological_diagnosis` VARCHAR(45) NULL DEFAULT NULL,
  `Variant_Interpretation` VARCHAR(50) NULL DEFAULT NULL,
  `dbSNP` VARCHAR(50) NULL DEFAULT NULL,
  `PopMax_AF` DECIMAL(19,9) NULL DEFAULT NULL,
  `gnomAD_exome_ALL` DECIMAL(19,9) NULL DEFAULT NULL,
  `ClinVar_ID` VARCHAR(50) NULL DEFAULT NULL,
  `ClinVar_SIG` VARCHAR(45) NULL,
  `ClinVar_DIS` VARCHAR(45) NULL,
  `COSMIC_ID` VARCHAR(45) NULL,
  `COSMIC_DIS` VARCHAR(45) NULL,
  `CADD_phred` VARCHAR(45) NULL,
  `ACMG_class_1` VARCHAR(20) NULL,
  `ACMG_class_2` VARCHAR(20) NULL,
  `ACMG_class_3` VARCHAR(20) NULL,
  `Chr` VARCHAR(5) NULL,
  `Start` INT NULL,
  `End` INT NULL,
  `Ref` VARCHAR(100) NULL,
  `Alt` VARCHAR(100) NULL,
  `Func.refGene` VARCHAR(45) NULL,
  `AF` VARCHAR(45) NULL,
  `AD` VARCHAR(45) NULL,
  `DP` INT NULL,
  `dnaSample_IdDna` VARCHAR(20) NOT NULL,
  `dnaSample_sample_IdSample` VARCHAR(20) NOT NULL,
  `dnaSample_sample_patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  PRIMARY KEY (`IdPatient`, `IdSample`, `Panel`, `dnaSample_IdDna`,
  `dnaSample_sample_IdSample`, `dnaSample_sample_patientInfo_IdPatient`),
  INDEX `fk_variantFiltred_dnaSample1_idx` (`dnaSample_IdDna` ASC,
  `dnaSample_sample_IdSample` ASC, `dnaSample_sample_patientInfo_IdPatient`
  ASC) VISIBLE,
  CONSTRAINT `fk_variantFiltred_dnaSample1`
    FOREIGN KEY (`dnaSample_IdDna`, `dnaSample_sample_IdSample`,
  `dnaSample_sample_patientInfo_IdPatient`)
    REFERENCES `bmf_db`.`dnaSample` (`IdDna`, `sample_IdSample`,
  `sample_patientInfo_IdPatient`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;
```

```
-----
-- Table `bmf_db`.`allVariants`
-----
```

```

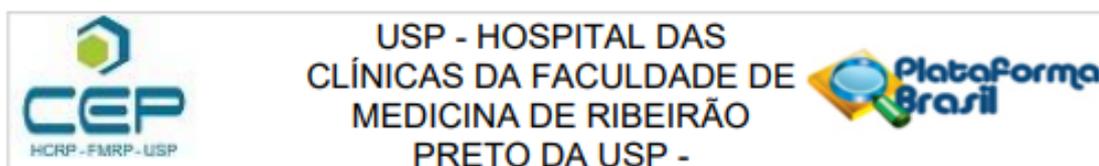
CREATE TABLE IF NOT EXISTS `bmf_db`.`allVariants` (
  `IdPatient` VARCHAR(10) NOT NULL,
  `IdSample` VARCHAR(20) NOT NULL,
  `Panel` INT NOT NULL,
  `Chr` VARCHAR(5) NULL,
  `Start` INT NULL,
  `End` INT NULL,
  `Ref` VARCHAR(100) NULL,
  `Alt` VARCHAR(100) NULL,
  `Gene.refGene` VARCHAR(45) NULL DEFAULT NULL,
  `Func.refGene` VARCHAR(10) NULL DEFAULT NULL,
  `ExonicFunc.refGene` VARCHAR(20) NULL DEFAULT NULL,
  `AChange.refGene` VARCHAR(200) NULL DEFAULT NULL,
  `dbSNP` VARCHAR(50) NULL DEFAULT NULL,
  `allele_count` INT NULL,
  `PopMax_AF` DECIMAL(19,9) NULL DEFAULT NULL,
  `ClinVar_ID` VARCHAR(50) NULL DEFAULT NULL,
  `ClinVar_SIG` VARCHAR(45) NULL,
  `ClinVar_DIS` VARCHAR(45) NULL,
  `COSMIC_DIS` VARCHAR(45) NULL,
  `COSMIC_ID` VARCHAR(45) NULL,
  `CADD_phred` VARCHAR(45) NULL,
  `AF` DECIMAL(19,9) NULL,
  `AD` DECIMAL(19,9) NULL,
  `DP` INT NULL,
  `gnomAD_exome_ALL` DECIMAL(19,9) NULL DEFAULT NULL,
  `gnomAD_genome_ALL` DECIMAL(19,9) NULL,
  `gnomAD_genome_AMR` DECIMAL(19,9) NULL,
  `X1000G_ALL` DECIMAL(19,9) NULL,
  `X1000G_AMR` DECIMAL(19,9) NULL,
  `ExAC_Freq` DECIMAL(19,9) NULL,
  `ExAC_AMR` DECIMAL(19,9) NULL,
  `ESP6500si_ALL` DECIMAL(19,9) NULL,
  `SIFT_pred` VARCHAR(45) NULL,
  `Polyphen2_HDIV_pred` VARCHAR(45) NULL,
  `LRT_pred` VARCHAR(45) NULL,
  `MutationTaster_pred` VARCHAR(45) NULL,
  `MutationAssessor_pred` VARCHAR(45) NULL,
  `FATHMM_pred` VARCHAR(45) NULL,
  `PROVEAN_pred` VARCHAR(45) NULL,
  `MetaSVM_pred` VARCHAR(45) NULL,
  `MetaLR_pred` VARCHAR(45) NULL,
  `M.CAP_pred` VARCHAR(45) NULL,
  `DANN_score` VARCHAR(45) NULL,
  `DANN_rankscore` VARCHAR(45) NULL,
  `fathmm.MKL_coding_pred` VARCHAR(45) NULL,
  `dnaSample_IdDna` VARCHAR(20) NOT NULL,
  `dnaSample_sample_IdSample` VARCHAR(20) NOT NULL,
  `dnaSample_sample_patientInfo_IdPatient` VARCHAR(10) NOT NULL,
  `dnaSample_IdSample` VARCHAR(20) NOT NULL,
  PRIMARY KEY (`IdPatient`, `Panel`, `IdSample`, `dnaSample_IdDna`,
  `dnaSample_sample_IdSample`, `dnaSample_sample_patientInfo_IdPatient`,
  `dnaSample_IdSample`),
  INDEX `fk_allVariants_dnaSample1_idx` (`dnaSample_IdDna` ASC,
  `dnaSample_sample_IdSample` ASC, `dnaSample_sample_patientInfo_IdPatient`
  ASC, `dnaSample_IdSample` ASC) VISIBLE,
  CONSTRAINT `fk_allVariants_dnaSample1`
  FOREIGN KEY (`dnaSample_IdDna`, `dnaSample_sample_IdSample`,
  `dnaSample_sample_patientInfo_IdPatient`, `dnaSample_IdSample`)
  REFERENCES `bmf_db`.`dnaSample` (`IdDna`, `sample_IdSample`,
  `sample_patientInfo_IdPatient`, `IdSample`)

```

```
        ON DELETE NO ACTION
        ON UPDATE NO ACTION)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8
COLLATE = utf8_unicode_ci;

SET SQL_MODE=@OLD_SQL_MODE;
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;
```


ANEXO 1 – Aprovação do Comitê de Ética em Pesquisa (CEP) em humanos



PARECER CONSUBSTANCIADO DO CEP

DADOS DA EMENDA

Título da Pesquisa: Correlação entre genótipo, comprimento telomérico e fenótipo nas telomeropatias

Pesquisador: Rodrigo do Tocantins Calado De Saloma Rodrigues

Área Temática:

Versão: 3

CAAE: 93617018.0.0000.5440

Instituição Proponente: Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da USP -

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 4.940.397

Apresentação do Projeto:

Conforme documento "PB_INFORMAÇÕES_BÁSICAS_1813826_E2.pdf":

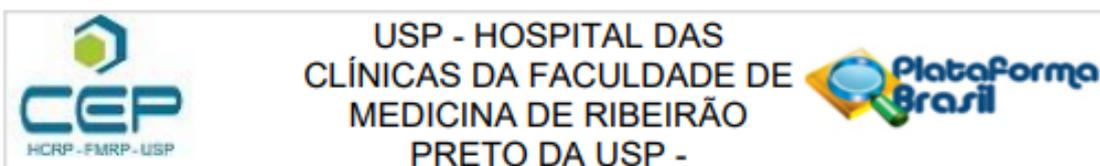
Justificativa da Emenda:

EMENDA SOLICITADA PARA ADEQUAÇÃO AO CEP DE INSTITUIÇÃO COPARTICIPANTE.

Resumo:

Telômeros são estruturas nucleoproteicas localizadas nas extremidades dos cromossomos lineares cuja função é manter a integridade genômica. Mutações em genes da telomerase ou em genes envolvidos na homeostase telomérica levam ao encurtamento excessivo dos telômeros, levando a senescência celular e/ou apoptose. Em humanos, a disfunção telomérica traduz-se no desenvolvimento de doenças envolvendo diferentes órgãos, como anemia aplástica (AA), disceratose congênita (DC), fibrose pulmonar idiopática (FPI), cirrose hepática e maior suscetibilidade ao câncer. Coletivamente, estas enfermidades ocasionadas pelo encurtamento telomérico excessivo são chamadas telomeropatias. Como a apresentação clínica é pleomórfica e a base genética, diversa, envolvendo diversos genes e diferentes padrões de herança, a correlação genótipo-genótipo é fundamental para a compreensão das vias celulares e moleculares de desenvolvimento de doença, assim como para determinar a relação entre genótipo

Endereço: CAMPUS UNIVERSITÁRIO
Bairro: MONTE ALEGRE **CEP:** 14.048-900
UF: SP **Município:** RIBEIRAO PRETO
Telefone: (16)3602-2228 **Fax:** (16)3633-1144 **E-mail:** cep@hcrp.usp.br



Continuação do Parecer: 4.940.397

e resposta às terapêuticas disponíveis. O primeiro objetivo deste projeto é identificar mutações em um painel de genes envolvidos na manutenção telomérica por sequenciamento de nova geração (NGS) em pacientes com diferentes manifestações de telomeropatias a fim de correlacionar o genótipo, comprimento telomérico, fenótipo e resposta ao tratamento. Adicionalmente, a expressão das RecQ helicases e TERRA (Telomeric-Repeat-Containing RNA), que pode contribuir como mecanismo molecular fisiopatológico, será avaliada em amostras de pacientes com doenças dos telômeros para identificar potenciais moduladores da regulação da replicação, reparo de dano ao DNA, integridade/manutenção telomérica e atividade da telomerase. As possíveis alterações nas vias das RecQ helicases e TERRA, desencadeadas por variações genéticas ou epigenéticas, serão avaliadas pela primeira vez para elucidação da ação destes componentes nas telomeropatias. Este projeto tem o potencial de elucidar as bases genéticas e moleculares envolvidas nas doenças dos telômeros, sua relação com a resposta ao tratamento e evolução clínica dos pacientes, contribuindo para o entendimento de mecanismos de doenças e descoberta de novos alvos terapêuticos.

Objetivo da Pesquisa:

Conforme documento "PB_INFORMAÇÕES_BÁSICAS_1813826_E2.pdf":

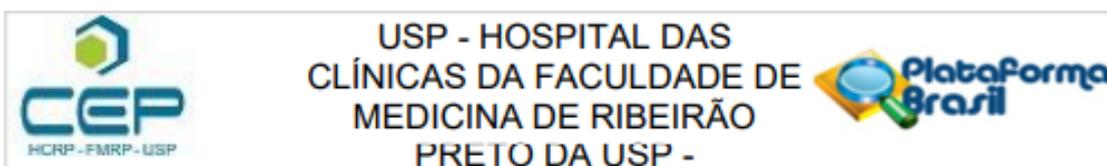
Objetivo Primário:

O presente estudo pretende identificar possíveis mutações nos genes associados às telomeropatias pela metodologia de Next-Generation Sequencing – NGS e targeting sequencing, bem como avaliar o perfil de expressão transcricional (mRNA) de genes da família RecQ helicase e de TERRA em células mononucleares do sangue periférico de pacientes que apresentem telômeros curtos com ou sem mutações na telomerase. Adicionalmente, o perfil de expressão destas helicases será relacionado com a intensidade do quadro clínico de falência medular e os transcritos de TERRA com o comprimento dos telômeros e o prognóstico dos pacientes com telomeropatias.

Objetivo Secundário:

- Determinar o comprimento telomérico dos pacientes deste estudo pelo método do Flow-FISH;
- Implementar a metodologia de NGS para um estudo genético dos pacientes com telomeropatias;
- Avaliar a expressão transcricional dos genes da família RecQ helicase (RECQL1, BLM, WRN, RECQL4 e RECQL5) em PBMCs dos pacientes alocados no estudo;
- Correlacionar os níveis da expressão dos genes da família RecQL helicases com a instabilidade

Endereço: CAMPUS UNIVERSITÁRIO
 Bairro: MONTE ALEGRE CEP: 14.048-900
 UF: SP Município: RIBEIRAO PRETO
 Telefone: (16)3602-2228 Fax: (16)3633-1144 E-mail: cep@hcrp.usp.br



Continuação do Parecer: 4.940.397

telomérica;

- Correlacionar os níveis da expressão dos genes da família RecQL helicases quanto ao desenvolvimento de AA e fibrose pulmonar idiopática ou familiar;
- Avaliar os níveis de expressão de TERRA nos pacientes do presente estudo;
- Correlacionar os níveis de expressão de TERRA com o comprimento telomérico dos pacientes e o prognóstico da doença.

Avaliação dos Riscos e Benefícios:

Conforme documento "PB_INFORMAÇÕES_BÁSICAS_1813826_E2.pdf":

Riscos:

Os riscos ou desconfortos da pesquisa são mínimos e estão relacionados à coleta da amostra de sangue, de pele, ao estoque temporário dessas amostras e de informações. Na coleta de sangue, o desconforto que você pode sentir é apenas da picada da agulha e um potencial hematoma (vermelhidão) que possa se formar no local da picada (cerca de 10% dos casos). Na "biópsia por punch", o procedimento é realizado com anestesia

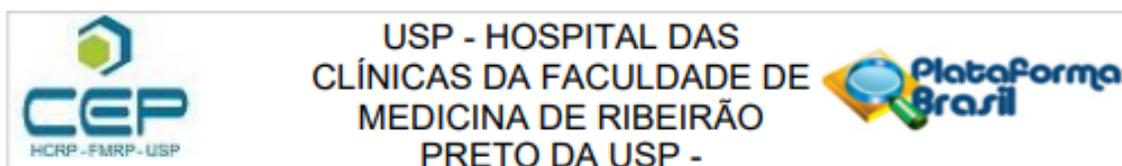
local e é necessário a sutura (dar um ponto). O desconforto que você pode sentir é mínimo e somente em decorrência do furo que se faz, como um beliscão. Infecção no local da coleta (de sangue e de pele) acontece muito raramente, pois o material utilizado em ambos os procedimentos é estéril e o local da picada ou da biópsia é bem higienizado.

Benefícios:

Ao identificar mutações em um painel de genes envolvidos na manutenção telomérica por sequenciamento de nova geração (NGS) em pacientes com diferentes manifestações de telomeropatias pode-se correlacionar o genótipo, comprimento telomérico, fenótipo e resposta ao tratamento. Além disso, a expressão das RecQ helicases e TERRA (Telomeric-Repeat-Containing RNA), podem contribuir com o entendimento do mecanismo

molecular fisiopatológico, em amostras de pacientes com doenças dos telômeros para identificar potenciais moduladores da regulação da replicação, reparo de dano ao DNA, integridade/manutenção telomérica e atividade da telomerase. As possíveis alterações nas vias das RecQ helicases e TERRA, desencadeadas por variações genéticas ou epigenéticas, serão avaliadas pela primeira vez para elucidação da ação destes componentes nas telomeropatias.

Endereço: CAMPUS UNIVERSITÁRIO
Bairro: MONTE ALEGRE **CEP:** 14.048-900
UF: SP **Município:** RIBEIRÃO PRETO
Telefone: (16)3602-2228 **Fax:** (16)3633-1144 **E-mail:** cep@hcrp.usp.br



Continuação do Parecer: 4.940.397

Comentários e Considerações sobre a Pesquisa:

Conforme documento "PB_INFORMAÇÕES_BÁSICAS_1813826_E2.pdf"

Justificativa da Emenda:

EMENDA SOLICITADA PARA ADEQUAÇÃO AO CEP DE INSTITUIÇÃO COPARTICIPANTE.

Considerações sobre os Termos de apresentação obrigatória:

Todos os termos da emenda foram devidamente apresentados ao CEP.

Recomendações:

Não há.

Conclusões ou Pendências e Lista de Inadequações:

Diante do exposto e à luz da Resolução CNS 466/2012, o CEP analisou e aprovou a EMENDA submetida em 23/08/2021, bem como o Projeto de Pesquisa versão 3 de 23 de agosto de 2021 e o Termo de Consentimento Livre e Esclarecido para Guarda de Material Biológico versão 3 de 23 de agosto de 2021.

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_1813826_E2.pdf	23/08/2021 14:03:46		Aceito
Outros	FORMENCAMINHAMENTO.pdf	23/08/2021 14:01:15	Rodrigo do Tocantins Calado De Saloma Rodrigues	Aceito
Outros	CARTAEMENDA2.PDF	23/08/2021 14:00:35	Rodrigo do Tocantins Calado De Saloma Rodrigues	Aceito
Projeto Detalhado / Brochura Investigador	ProjetoRTCCEPversao3.pdf	23/08/2021 14:00:13	Rodrigo do Tocantins Calado De Saloma Rodrigues	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_PACIENTE_LAB_RTC_v3.pdf	23/08/2021 13:59:15	Rodrigo do Tocantins Calado De Saloma Rodrigues	Aceito

Endereço: CAMPUS UNIVERSITÁRIO

Bairro: MONTE ALEGRE

CEP: 14.048-900

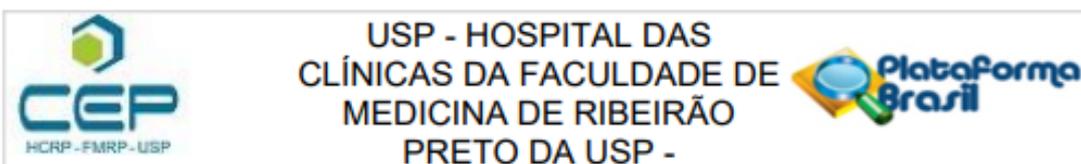
UF: SP

Município: RIBEIRAO PRETO

Telefone: (16)3602-2228

Fax: (16)3633-1144

E-mail: cep@hcrp.usp.br



Continuação do Parecer: 4.940.397

Outros	Solic_criac_bior.PDF	21/05/2021 11:36:20	Rodrigo do Tocantins Calado De Saloma Rodrigues	Aceito
Cronograma	CronogramaEMENDA.docx	06/04/2021 09:26:32	Rodrigo do Tocantins Calado De Saloma Rodrigues	Aceito
Outros	RelatorioCEP2021.docx	06/04/2021 09:15:31	Rodrigo do Tocantins Calado De Saloma Rodrigues	Aceito
Declaração de concordância	CARTAEMENDA.pdf	06/04/2021 09:15:00	Rodrigo do Tocantins Calado De Saloma Rodrigues	Aceito
Folha de Rosto	Folharosto.PDF	14/06/2018 08:54:22	Rodrigo do Tocantins Calado De Saloma Rodrigues	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

RIBEIRAO PRETO, 30 de Agosto de 2021

Assinado por:
MARCIA GUIMARÃES VILLANOVA
(Coordenador(a))

Endereço: CAMPUS UNIVERSITÁRIO
Bairro: MONTE ALEGRE CEP: 14.048-900
UF: SP Município: RIBEIRAO PRETO
Telefone: (16)3602-2228 Fax: (16)3633-1144 E-mail: cep@hcrp.usp.br