UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
PROGRAMA DE PÓS-GRADUAÇÃO EM ONCOLOGIA CLÍNICA, CÉLULAS-TRONCO E TERAPIA CELULAR

FERNANDA GUTIERREZ RODRIGUES

Identificação de moduladores genéticos em pacientes com anemia aplástica por sequenciamento de nova geração

Genetic screening of patients with aplastic anemia by targeting sequencing

Ribeirão Preto – São Paulo
2017

FERNANDA GUTIERREZ RODRIGUES

Identificação de moduladores genéticos em pacientes com anemia aplástica por sequenciamento de nova geração

Genetic screening of patients with aplastic anemia by targeting sequencing

Versão original

Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Doutor em Ciênicas.

Programa de Oncologia Clínica, Células-Tronco, e Terapia Celular. Área de Concentração: Diferenciação Celular Normal e Neoplásica

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

Ribeirão Preto

2017

FICHA CATALOGRÁFICA

GUTIERREZ-RODRIGUES, Fernanda

Identificação de moduladores genéticos em pacientes com anemia aplástica por sequenciamento de nova geração – Ribeirão Preto, 2017.

108 p. : il. ; 30 cm

Dissertação de Mestrado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Programa de Oncologia Clínica, Células-Tronco e Terapia Celular. Área de Concentração: Diferenciação Celular Normal e Neoplásica.

Orientador: Calado, Rodrigo Tocantins

1. Anemia Aplástica. 2. Sequenciamento de nova geração. 3. *RTEL1*. 4.Telômeros e telomerase.

Nome: GUTIERREZ-RODRIGUES, Fernanda

Título: Identificação de moduladores genéticos em pacientes com anemia aplástica por sequenciamento de nova geração

Title: Genetic screening of patients with aplastic anemia by targeting sequencing

Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Doutor em Ciências, Programa de Oncologia Clínica, Células-Tronco e Terapia Celular; Área de Concentração: Diferenciação Celular Normal e Neoplásica.

Aprovado em: ___/___/_____

Banca Examinadora

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

Julgamento: _____ 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:

- Fundação de Amparo à Pesquisa do Estado de São Paulo – **FAPESP**
  - Bolsa de doutorado no país: **2014/27294-7**
    Período: 01/05/2015 a 16/011/2017
  - Bolsa Estágio de Pesquisa no Exterior (BEPE): **2015/18491-6**
    Período: 11/06/2016 a 10/01/2017
  - Centro de Terapia Celular – CTC (CEPID): **2013/08135-2**
- Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (**FMRP** – USP)
- Fundação Hemocentro de Ribeirão Preto (**FUNDHERP** – HCFMRP – USP)
- 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** – HCFMRP – USP)
- National Institutes of Health – **NIH** (Bethesda, MD, USA)

We can make our plans, but the LORD determines our steps.

Proverbs 16:9 (NLT)

AGRADECIMENTOS / *ACKNOWLEDGMENTS*

Ao meu orientador **Prof. Dr. Rodrigo T. Calado** pelo apoio e oportunidade de permenacer em seu laboratório no doutorado. Muita obrigada pela confiança depositada no meu trabalho, pela liberdade em conduzir nossa pesquisa, por me incentivar a dar o meu melhor e por permitir que eu voasse para terras mais distantes. Você é parte fundamental nesta jornada profissional.

My sincere gratitude to **Dr Neal S. Young** for the time that I spent in his lab in 2016. Thank you for allowing me to be creative and independent but also to be part of this amazing team in your lab. You are an inspiration as a mentor, doctor, and a person.

I would like to thank **Dr Danielle Townsley** for the doors she opened for me and to allow me to be part of different projects that shaped my understating of science. Thank you for pushing me when I needed and for the time you spent teaching me. You are part of my journey.

To **Dr James Cooper,** for the friendship and all the times that he mentored me. Thank you for believing that I was capable of going beyond in my research and to show me the perspectives of the scientific world.

I don't have enough words to express my gratitude to **Dr Sachiko Kajigaya**. Thank you for the endless time you spent explaining me the protocols or reviewing my never-ending paper. Thank you for caring about me, physically and mentally. Thank you for all the time you stopped by and gave me food. Thank you for being part of my every day in the lab and to shape me into what I am now as a scientist.

Aos meus colegas de trabalho e queridos amigos do Telomelab, muito obrigada pela paciência e apoio nos momentos difíceis. Todos vocês são muito queridos! Em especial, à **Flávia e ao André.** O que eu tenho não é suficiente para retribuir tudo o que vocês representam para mim. Ao **João Paulo**, pela companhia e amizade sincera. **À Raquel,** pela amizade, pelo respeito, pelos puxões de orelha e pela colaboração. Você é o meu exemplo mulher independente e cientista competente.

I thank my entire fellows in NHLBI. Special thanks to **Pilar, Valentina, Alice, Anna, and Lauren** for the friendship and to make my days happier. Dumplings, pizzas,

and eggplant parmigiana. Thank you for helping in my experiments, to worry and be happy about my results and to let me be part of your time at NIH.

À Katarina, Renata, Lemão, Jaque, Ana, Fer, Lilian, Dalvinha, Juan, Diego, Virginia, Laura, Kamila, Nayana, Anne, Larissa, por me presentear com um carinho, amor e amizade que eu não merecia.

À Marol, Jose, Moara, Leandro, Marina, e Renato, por me acolherem e serem parte da minha familia em Bethesda.

To Venina, Luciana, Zach, Ken, Alina, Scott, and all Citizen Heights, for being my family. You are the perfect translation of God's love for me.

Aos meus pais, meu irmão e minha familía, por simplesmente serem o que vocês são na minha vida. Vocês são o meu tesouro. Amo vocês.

A Deus, o maior merecedor de toda minha gratidão. Obrigada pela graça, pelo amor incondicional e por cuidar de cada detalhe dos meus dias passados, presentes e dos futuros.

RESUMO

GUTIERREZ-RODRIGUES, F. **Identificação de moduladores genéticos em pacientes com Anemia Aplástica por Sequenciamento de nova geração.** 2017. 108 f. Tese (Doutorado) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, São Paulo, 2017.

A fisiopatologia das síndromes de falência da medula óssea (FMO) está relacionada a mecanismos adquiridos de destruição das células-tronco hematopoeiticas na medula ou a defeitos constitucionais em genes fundamentais para o reparo do DNA e manutenção dos telômeros. A anemia aplástica (AA), o protótipo das doenças de FMO, pode ter etiologia adquirida ou constitucional. A avaliação genética de pacientes com AA adquirida tem como objetivo a detecção de mutações somáticas que possam ser usadas como marcadores de resposta ao tratamento imunossupressor. Diferentemente, em pacientes com AA constitucional, a avaliação genética é fundamental para detecção de mutações etiológicas na doença do paciente, sendo essencial para o tratamento e seleção de doadores de medula óssea. Contudo, o papel das mutações constitucionais na fisiopatologia e modulação imunológica da AA adquirida ainda não é conhecido. Neste estudo, nós sequenciamos pacientes com AA de duas coortes independentes utilizando diferentes painéis de sequenciamento de genes alvos. A primeira coorte, composta por 13 pacientes com AA adquirida, foi sequenciada utilizando um painel com 165 genes relacionados à FMO, neoplasias hematológicas, reparo de DNA, manutenção dos telômeros e vias de resposta imune. A segunda coorte, composta por 59 pacientes investigados para doença constitucional, foi sequenciada com um painel de sequenciamento comercial com 49 genes relacionados à FMO hereditária. Foram identificadas alterações potencialmente patogênicas em três dos cinco pacientes com AA adquirida que não responderam à imunossupressão: dois pacientes com variantes em *TERT* e um com uma variante em *DHX36.* Não foram identificadas variantes funcionalmente relevantes nos pacientes que responderam ao tratamento imunossupressor. Em contraste, foram identificadas variantes potencialmente patogênicas em *RTEL1* em 8 pacientes com AA constitucional. Variantes em *RTEL1* foram associadas tanto ao encurtamento telomérico quanto à erosão excessiva do 3' *overhang*, independentemente do comprimento dos telômeros. Desse modo, apenas a medida do comprimento dos telômeros não foi suficiente para identificar todos os

pacientes com disfunções teloméricas. As plataformas de sequenciamento de nova geração diminuíram o custo e o tempo para a avaliação genética dos pacientes com FMO. Em nosso estudo, os pacientes com AA adquirida não apresentaram um padrão genético associado à sua resposta ao tratamento com imunossupressores, no entanto, o sequenciamento da coorte com suspeita de AA constitucional foi capaz de identificar o defeito genético associado à doença do paciente em 40% dos casos. O uso de dados clínicos, investigação familiar, análises *in silico* e testes funcionais foram essenciais para uma correta interpretação da patogenicidade de novas variantes identificadas por sequenciamento de nova geração.

Palavras-chave: anemia aplástica, sequenciamento de nova geração, *RTEL1*, disfunção telomérica.

ABSTRACT

GUTIERREZ-RODRIGUES, F. **Genetic screening of patients with aplastic anemia by targeting sequencing.** 2017. 108 f. Tese (Doutorado) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, São Paulo, 2017.

The pathophysiology of bone marrow failure (BMF) can be immune, as in acquired aplastic anemia (AA), or constitutional, due to germline mutations in genes critical for DNA repair and telomere maintenance. The genetic screening of patients with constitutional AA is performed to detect germline mutations that are etiologic in patients' disease. That is critical for treatment decisions and to identify a donor for a bone marrow transplant. In acquired AA, the genetic screening has been used to detect somatic mutations that can predict patients' outcomes after treatment, as the role of germline mutations in this disease is yet not clear. To investigate the role of germline variants in AA, we screened two independent cohorts with two different targeting sequencing panels; a first cohort composed by 13 patients with acquired AA that was screened using a panel with 165 genes related to BMF, hematologic malignancies, DNA repair, telomere maintenance, and immune response pathways. A second cohort composed of 59 patients suspected to have a constitutional disease screened by a commercial Inherited Bone Marrow Failure Sequencing panel. In our first cohort, while patients without functional relevant germline variants responded to immunosuppression treatment (n=8), three out of 5 nonresponder patients were identified with variants in telomere biology genes. We found patients carrying *TERT* and *DHX36* variants. In our constitutional AA cohort, we identified 8 patients carrying variants in the *RTEL1* gene, a helicase critical to telomere maintenance. *RTEL1* variants associated with both patients' overall telomere shortening and single-stranded 3' overhang erosion independent of telomere length. Also, 3' overhang erosion was associated with patients' predisposition to clonal evolution. In this context, the variants identified in the helicases genes *DHX36* and *RTEL1* were both associated with patients' normal telomere length and poor outcomes. Also, telomere length measurement alone was insufficient to identify all primary telomere defects. The platforms of next-generation sequencing decreased the cost and time for the genetic screening of patients with BMF. In our study, acquired AA patients did not display a clear genetic pattern associated with their immunosuppressive treatment response. In contrast, the sequencing of the cohort selected based on their suspicion to have an inherited disease

identified a molecular defect that might be pathogenic in up to 40% of patients, including the *RTEL1* variants. Pathogenicity assessment of genetic variants requires a combination of clinical, *in silico*, and functional data required to avoid misinterpretation of common variants.

Keywords: aplastic anemia, next-generation sequencing, *RTEL1*, telomere dysfunction

Figures

Tables

Table of Contents

INTRODUCTION

1. INTRODUCTION

1.1. Bone marrow failure syndromes

Bone marrow failure (BMF) syndromes are generally characterized by bone marrow aplasia and peripheral blood pancytopenia. Aplastic anemia (AA) is the main example of BMF caused by either an acquired immune destruction of hematopoietic stem and progenitor cells (Young *et al.*, 2006), or genetic defects associated to inherited diseases at any stage of life. In most patients with AA the destruction of hematopoiesis is due to an immune mechanism dominated by activated cytotoxic T cells that target the hematopoietic progenitors in the bone marrow (BM) (Young and Maciejewski, 1997; Young *et al.*, 2006; Young *et al.*, 2008; Calado and Young, 2009). After exposure to an initiating antigen, cells and cytokines of the immune system act destructively on stem cells in the marrow, reducing their pool and the normal levels of circulating leukocytes, erythrocytes, and platelets in peripheral blood. Activated cytotoxic T lymphocytes have a major role in tissue damage by secreting interferon-γ (IFN-γ) and tumor necrosis factor (TNF). Both activate interferon regulatory factor 1 (IRF-1) signaling pathways that inhibits the overall transcription and cell cycle of hematopoietic stem cell (HSC) (Figure 1). Destruction of HSC are also caused by apoptosis induced when the Fas receptor is activated by it ligand, a cytokine produced by cytotoxic T lymphocytes. The mechanism is supported by IFN-γ and TNF secretion, as both upregulate the expression of Fas receptor in the HSC, and an increased production of interleukin-2, that leads to polyclonal expansion of T cells.

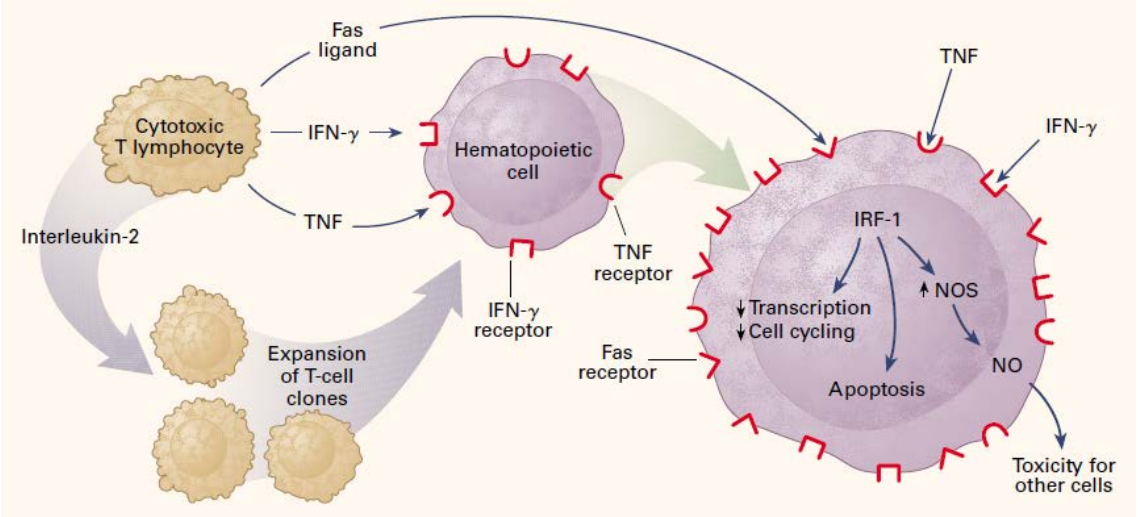


Figure 1. The pathophysiology of acquired aplastic anemia. Immune destruction of hematopoiesis drive by activated cytotoxic T cells that target the hematopoietic progenitors in bone marrow (Young and Maciejewski, 1997). Reproduced under authors' consent.

A subgroup of AA patients carries mutations in telomere and telomerase related genes and develop the disease due to a hematopoietic stem cell pool depletion caused by a telomere dysfunction (Calado and Young, 2009; Townsley *et al.*, 2014). In addition to AA, telomere defects are associated with a spectrum of disorders collectively called telomeropathies, such as dyskeratosis congenita (DC), liver cirrhosis, and idiopathic pulmonary fibrosis (IPF) (Calado and Young, 2009).

The telomeres are consisted of a double-stranded DNA with thousands of *in tandem* TTAGGG repeats in it sequence, associated with a group of at least six proteins named shelterins (TRF1, TRF2, RAP1, POT1, TIN2, and TPP1). This DNA-protein complex caps the chromosome ends and avoids unappropriated DNA machinery repair activation (Figure 2A). At the telomere terminus, there is a short 3' single-stranded G-rich overhang that invades the telomeric double helix to form a circle structure, the t-loop (Figure 2B). The t-loops structurally avoid recognition of chromosome's free ends as double-stranded breaks by DNA repair machinery and functions to prevent inappropriate DNA damage

response (DDR) and genome instability (Moyzis *et al.*, 1988; Palm and De Lange, 2008). The loss of telomeric repeats after cell division is a physiological consequence of aging, as DNA polymerase is unable to fully duplicate telomeres (Collins and Mitchell, 2002). When critically short, telomeres signal cells to senescence and apoptosis via activation of the p53/p21 pathways. Erosion of 3' overhang is also damaging to the telomere structure, as it leads to replicative senescence and collapse of t-loop in cells (Stewart *et al.*, 2003). In highly proliferative cells, telomere length (TL) is maintained by telomerase, an enzyme composed of the catalytic subunit telomerase reverse transcriptase (TERT), an RNA template (TERC), and associated proteins (Figure 2A).

The telomeropathies, or telomere diseases, are a group of disorders caused by germline mutations in telomere-associated genes, resulting in excessive telomere attrition, reduced stem cell pool and regenerative potential, and clinically affecting the BM, lung, liver, and skin among other tissues (Figure 2A, 2C) (Calado and Young, 2009; Paiva and Calado, 2014). Inheritance of telomere diseases can be autosomal dominant, autosomal recessive, or X-linked, depending on the gene mutated and the penetrance of the mutation (Knight *et al.*, 1999; Vulliamy *et al.*, 2005; Yamaguchi *et al.*, 2005; Calado and Young, 2008; Savage *et al.*, 2008; Collopy *et al.*, 2015). To date, mutations in the genes *DKC1, TERC, TERT, USB1, CTC1, NOP10, NHP2, WRAP53, TINF2, RTEL1, PARN,* and *ACD* have been identified in patients with the spectrum of telomere diseases (Knight *et al.*, 1999; Vulliamy *et al.*, 2005; Marrone *et al.*, 2007; Vulliamy *et al.*, 2008; Walne *et al.*, 2008; Calado and Young, 2009; Anderson *et al.*, 2012; Walne *et al.*, 2013; Tummala *et al.*, 2015). Additionally, telomerase mutations and short telomeres are genetic risk factors for the development of some hematologic cancers, including the MDS, acute myeloid leukemia (AML), and chronic lymphocytic leukemia (CLL) (Paiva and Calado, 2014).

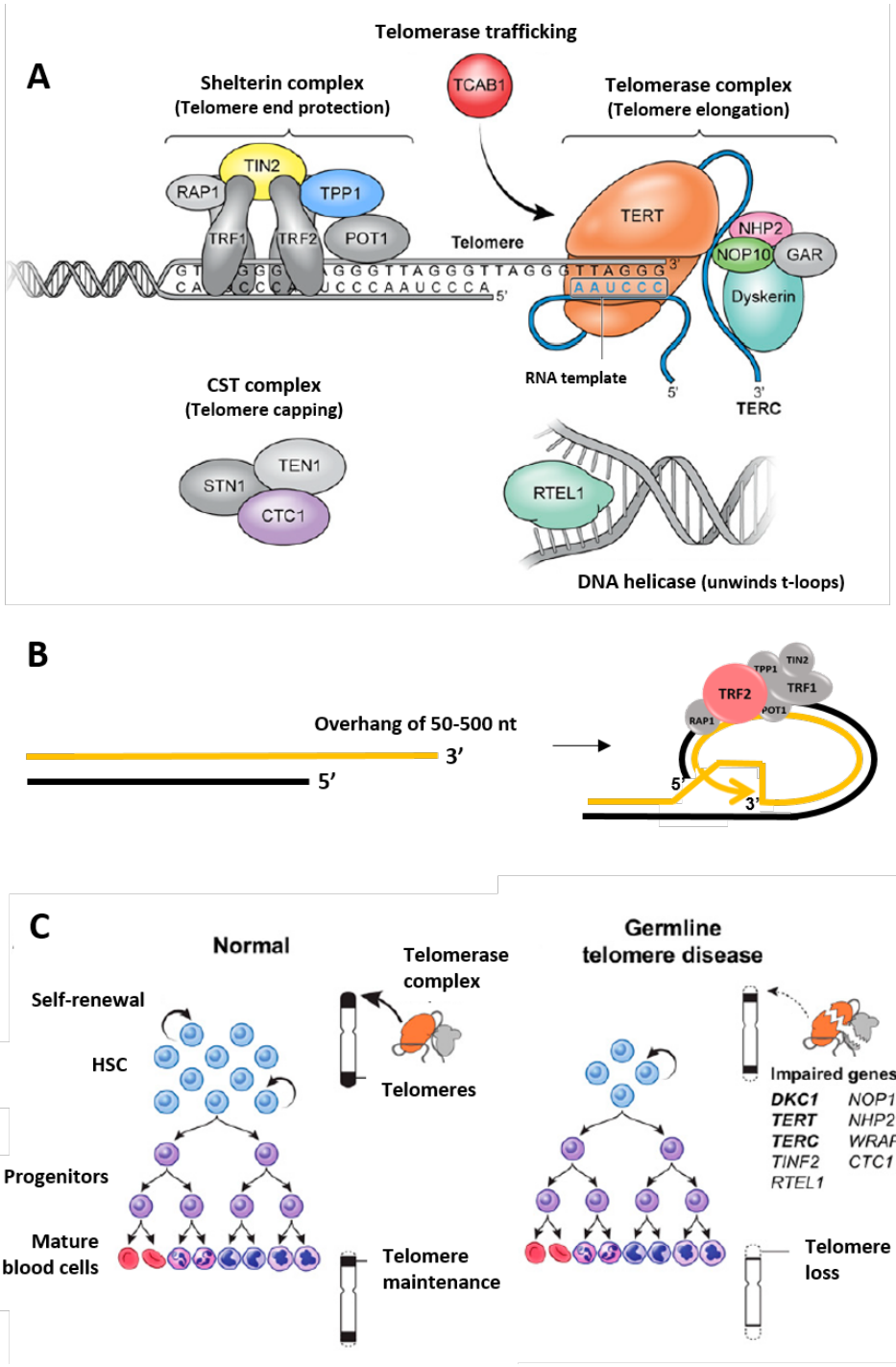


Figure 2. Telomere and telomerase biology. (A) Components involved in telomere maintenance associated with telomere diseases. Mutations in the genes DKC1 (encodes the dyskerin protein), TERC, TERT, USB1, CTC1, NOP10, NHP2, WRAP53 (encodes the TCAB1 protein), TINF2 (encodes the TIN2 protein), RTEL1, PARN, and ACD, have been linked to telomere diseases as they impair the mechanisms of telomere protection, elongation and stability. This figure was kindly provided by Dr Danielle Townsley. **(B) The end of telomeres.** The end of telomeres in the 3' strand is longer than the 5' strand. This protrusion creates a single-stranded G-rich overhang with 50-500 nucleotides (nt) that invades the telomeric double helix to form a circle structure, the t-loop. **(C) Mechanism of telomere attrition.** In normal individuals, telomeres shorten in

hematopoietic cells (HSC) with each cell division. Telomerase counterbalance telomere attrition by adding TTAGGG repeats at telomeres ends as an attempt to maintain hematopoiesis. Germline mutations in the telomere biology genes lead to a reduced stem cell pool due to severe telomere loss during replication. This figure was kindly provided by Dr Danielle Townsley.

As telomerase, RTEL1 (regulator of telomere elongation) is critical for telomere maintenance and DNA repair. RTEL1 is a helicase with a variety of functions related to genome integrity and DNA replication: (1) disassembly of DNA secondary structures, such as t-loops and G4 quadruplexes, to promote telomere replication and elongation by telomerase (Vannier *et al.*, 2012); (2) regulation of homologous recombination in cells (Uringa *et al.*, 2011); and (3) involvement in nuclear and cytoplasmic trafficking of ribonucleotide proteins (Schertzer *et al.*, 2015). T-loop disassemble activity is dependent of TRF2 and RTEL1 interaction through a specific RTEL1 domain, the C4C4-RING domain (Sarek *et al.*, 2015). The C4C4-RING is a highly conserved domain characterized by the Cys–X2–Cys–X9–Cys–X2–Cys–X4–Cys–X2–Cys–X10–Cys–X2–Cys sequence. This metal coordinate motif is only present in RTEL1 isoform 3 (1300 aa; NCBI: NM_001283009.1 and NP_001269938.1; Ensembl: ENST00000360203.9). Despite, RTEL1 isoform 2 (1243 amino acids, NCBI NM_032957.4) be transcribed in similar levels than isoform 3 in human cells, isoform 3 stands as important for t-loop dismantle and telomere replication. Even though shorter, isoform 2 has the other major RTEL1 domains (RAD3, Harmonin-like and PIP domains) (Figure 3).

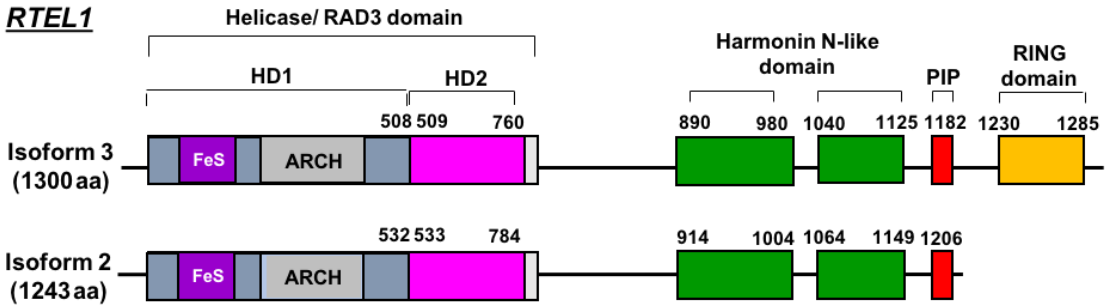


Figure 3. Linear representation of *RTEL1* gene. Linear representation of RTEL1 isoforms 3 (1300 amino acids; NCBI NM_016434.3) and 2 (1243 amino acids, NCBI NM_032957.4). The heterogeneity of RTEL1 functions is attributed to its different domains: RAD3, the catalytic core responsible for DNA unwinding activity; PIP, required for binding to proliferating cell nuclear antigen (PCNA) and promotion of G4 quadruplexes unwinding; C4C4-RING, which resolves t-loops during telomere replication; and the Harmonin N-like, a region of unclear function.

Biallelic mutations in *RTEL1*, either in homozygosity or compound heterozygosity, provoke extreme telomere erosion and severe phenotype, clinically manifested as DC and Hoyeraal-Hreidarsson (HH) syndrome in early childhood (Ballew *et al.*, 2013; Deng *et al.*, 2013; Le Guen *et al.*, 2013; Walne *et al.*, 2013; Touzot and Kermasson, 2016). TERT or TERC haploinsufficiency due to heterozygous mutations also results in telomere shortening and is associated with less severe phenotype and is identified in up to 10% of patients with AA or myeloid malignancies (Yamaguchi *et al.*, 2003; Yamaguchi *et al.*, 2005; Calado, Regal, Hills, *et al.*, 2009). Heterozygous *RTEL1* mutations have been reported in idiopathic pulmonary fibrosis (IPF) (Kannengiesser *et al.*, 2015; Stuart *et al.*, 2015), but is not common in hematologic disorders.

1.2. Treatments for acquired vs. constitutional aplastic anemia

Hematopoietic stem cell transplantation (HSCT) or non-transplant therapies such as immunosuppressive regimen (IST) can reverse the process of BMF, leading to improved hematopoiesis and partial or full resolution of pancytopenia (Young and Maciejewski,

1997). Non-transplant therapies for both acquired and constitutional BMF have expanded recently and in several cases, are the first line of treatment. Historically, the most successful IST regimens have combined cyclosporine and antithymocyte globulin (ATG), which is a polyclonal antibody produced in horse or rabbit inoculated with human thymocytes (Young *et al.*, 2006; Scheinberg *et al.*, 2011). The IST induced by ATG proved to be effective, with long-term survival, and with results similar to the stem-cell transplantation from a histocompatible sibling; the addition of cyclosporine to ATG regimen improved the response rate and survival of the AA patients (Scheinberg *et al.*, 2011). The responsiveness of AA to IST is the best evidence of an immune pathophysiology mechanism of this disease (Young *et al.*, 2006). However, germline mutations do not excluded a coexisting immune component in the inherited marrow failure, as some patients with telomere gene mutations and AA/BMF may respond to immunosuppressive therapy (Townsley *et al.*, 2014).

The purpose of IST in acquired AA is to restore the bone marrow function, to recovery the hematologic counts, to prevent relapses, and to improve the life expectancy of patients. IST is recommended for (1) patients with non-severe AA who are transfusion dependent, (2) patients with severe or very severe disease who are >40 years old, and (3) younger patients with severe or very severe disease who do not have an HLA-identical sibling donor (Marsh *et al.*, 2009; Scheinberg and Young, 2012).

IST response is dependent on ATG preparation used for treatment. About 60% of patients are responders at 3 or 6 months after initiation of horse ATG, while only 30% of patients are responders after rabbit ATG treatment at the same time (Scheinberg *et al.*, 2011). In the absence of histocompatible sibling donors, a second course of rabbit ATG is administered to recover patients failing to respond to an initial course of ATG, with a response rate around 80%. It is also expected that one third of responders will relapse

after initial therapy with horse ATG, in which is necessary either a reintroduction or increasing of cyclosporine, or a repeated IST course. The rate of non-responders can be higher when rabbit IST is administered as a first-line treatment (Young *et al.*, 2006; Scheinberg *et al.*, 2011).

These differences in the IST response are also seen in our "Brazilian Marrow Failure Network" (Clé *et al.*, 2015), which reinforces the hypothesis that the mechanisms by which the horse and rabbit preparations restore BM function are variable. In Brazil, the rate of non-responders could reach 55% of patients as they are treated with rabbit ATG because horse ATG or eltrombopag are not commercially available (Clé *et al.*, 2015). Both ATG preparations deplete cytotoxic T cells, however a more prolonged lymphopenia and a profound CD4+ depletion is observed after rabbit ATG administration. These findings suggest that use of rabbit ATG might impair marrow recovery by the depletion of stimulatory cytokines derived from T cells (Marsh *et al.*, 2009; Scheinberg and Young, 2012). In general, independent of ATG preparation, one third of AA patients will be responders to IST while one third will be refractory, not recovering from pancytopenia even 6 months after IST. In these cases, because of the lack of histocompatible sibling donors, a second course of rabbit ATG is used to salvage patients failing to respond to an initial course of ATG, with a response rate around 80%. A third course of IST in refractory AA is not recommended (Marsh and Kulasekararaj, 2013; Young, 2013). The mechanisms by which some patients persist with severe pancytopenia remain unknown.

The hypotheses put forward to explain the lack of IST responsiveness after a first course of ATG are: (1) a constitutional AA and/or telomere disease is the underlie mechanism of patients' disease instead of a immune mediated mechanism, (2) a low number of residual stem cell in bone marrow capable of reconstitute the hematopoiesis,

(3) an inadequate immunosuppression regimen or the use of rabbit ATG in a first IST course (Calado and Young, 2008; Scheinberg and Young, 2012; Marsh and Kulasekararaj, 2013). All these factors may modulate IST response and may explain the response differences seen in AA patients front of the same ATG + cyclosporine IST.

Treatment of patients with refractory AA has been challenging, often resulting in referral to transplant. Recommendations for initial allogeneic stem cell transplantation *vs.* IST has depended on age, comorbidities, blood count severity, and donor availability (Townsley and Winkler, 2016). Even after the exhaustion of available therapies, approximately 20% of AA patients do not have an option of treatment if an HLA-identical sibling donor is not available.

In constitutional AA, the HSCT is the only curable treatment. As for refractory patients with acquired AA, the identification of a compatible donor is a limiting factor that often leads patients to non-transplanting therapies to manage their disease. For patients with telomeropathies, the androgen treatment led to telomere elongation and hematologic responses in 19 of 24 patients (79%) who could be evaluated at 3 months and in 10 of 12 patients (83%) who could be evaluated at 24 months (Townsley *et al.*, 2016). At the National Institutes of Health, clinical trials have been conducted to improve the response rate of AA patients that underwent nontransplant therapies. Recently, trilineage hematologic responses could be achieved with the eltrombopag in refractory AA after years of failed trials using other growth factor agents. Eltrombopag is a small-molecule oral thrombopoietin receptor agonist that can promote HSC expansion in bone marrow (Townsley and Winkler, 2016). Dr. Neal S. Young and Dr. Danielle M. Townsley have been introduced eltrombopag (GSK/Novartis Pharmaceuticals) as a first line therapy for AA in combination with ATG and cyclosporine (conventional IST treatment). The addition of eltrombopag increased AA patients' complete response rate from 70% to 95%

after 6 months of treatment (Townsley *et al.*, 2017). Also, in a phase 2 study conducted by Dr. Cynthia Dunbar and Dr. Thomas Winkler, 40% of patients with severe AA refractory to at least 1 course of standard IST responded at 3 to 4 months to eltrombopag salvage therapy, including tri- and bilineage responses (Desmond *et al.*, 2014). Despite these remarkable results, the mechanisms by which some patients persist with severe pancytopenia after standard IST are unclear. Also, further studies are required to evaluate if eltrombopag treatment increases the risk of clonality in hematopoietic cells and contributes to the development of malignant diseases.

In the context of an acquired mechanism associated with AA, we hypothesized that genetic factors might modulate the immune response in these patients and play a role in patients' response to IST and outcomes. In constitutional AA, germline mutations are etiologic in patients' disease. The identification of genetic defects that associate with the disease and the mechanism by which it perturbs the cells homeostasis are the challenges in patient's treatment.

1.3. Genetic screening of patients with aplastic anemia

In many clinical services, BMF patients are consecutive screened for TL to distinguish the etiology of their disease as acquired or constitutional. Most treatment algorithms propose to measure TL of all patients with suspected constitutional BMF in peripheral blood and to screen relatives in affected families (Calado and Young, 2008). In acquired AA, telomere dysfunction is not etiologic, but patients' short telomeres may be an evidence of a history of the HSC replicative stress instead of an underlying constitutional defect.

Several methods are available to measure TL, but the majority of studies and diagnostic laboratories use one of the following methods: (1) terminal restriction

fragment (TRF) analysis by Southern blot, (2) fluorescence in situ hybridization combined with flow cytometry (flow-FISH), or (3) quantitative PCR (qPCR) (Gutierrez-Rodrigues *et al.*, 2014). For each technique, the TL of healthy individuals across different age ranging are measured and plotted according to age. Distribution curves are derived from best-fit analysis of TL to adjust the 1st, 10th, 50th, 90th, 99th percentiles to the curve. Very short telomeres are defined as below the first percentile (1st) for age-matched controls and short telomeres as below the tenth (10th) percentile.

The genetic screening of BMF patients for germline mutations in BMF-related genes is also critical to treatment decisions and should be performed along with TL measurement in cases suggestive of constitutional BMF. Germline mutations are etiologic not only in telomeropathies but also in Fanconi anemia, Blackfan-Diamond syndrome, Shwachman-Diamond syndrome, and congenital neutropenia among others (Khincha and Savage, 2013). In the BMF clinic at FMRP/USP/Brazil, we screen all patients for telomere defects using the TL measurement, but only the highly suspected cases of constitutional diseases are screened for germline mutations using the Sanger sequencing method.

The normal TL and the diagnosis of acquired AA do not diminish the importance to genetic screen AA patients. For decades, the acquired AA was known to be driven by an immune-mediated mechanism rather than a malignant disease caused by somatic or germline mutations in the myeloid compartment. However, it was reported that clonal hematopoiesis driven by somatic mutations in a particular group of genes are common in acquired AA and can be associated with response, survival, and development of MDS or AML (Yoshizato *et al.*, 2015). This was the first evidence that acquired genetic alterations present in a subset of cells that clonally expanded play a role in patients' disease and are associated with malignant hematologic disorders. But these findings are still controversial

across studies. Recently, a clinical trial did not find a correlation between the AA patients' clonal mutational status and response to IST/Eltrombopag (Townsley *et al.*, 2017). The only factor associated with patients' better outcome was TL higher than the 10^{th} percentile of age-matched controls, hypothesized to represent the replicative potential of the hematopoietic stem cell in BM.

In general, the genetic screening of patients with constitutional AA is performed to detect germline mutations that are the cause of patients' disease, which is critical for treatment decisions and BM donor selection. In acquired AA, the genetic screening has been used to detect somatic mutations that can predict patients' outcomes as the role of germline mutations in this disease is not clear.

The next-generation sequencing (NGS) technologies may represent a powerful tool for screening of AA patients. The NGS can determine the entire human genome sequence, by the whole-genome sequencing (WGS), or only the coding sequences (exome), by the whole-exome sequencing (WES). In both methodologies, the entire genome is broken into small fragments and attached to special adapters that will bind to complementary oligos immobilized on the surface of a flow-cell to sequencing. With the NGS, millions of these DNA segments from across the genome are sequenced at the same time. NGS devices employ this massively parallel sequencing to deliver very inexpensive sequence data. The sequence regions are then compared to a reference human genome (Bick and Dimmock, 2011). These new methods are becoming central in human disease research and are starting to be used in the routine of clinical care. The WGS/WES are very useful to fully understand the genome variation of human diseases, as well as the patients' genetic susceptibility to treatment (Pareek *et al.*, 2011). The development of NGS has dramatically reduced the cost and increased the output of sequencing since the Sanger method remains costly and time-consuming.

This project was first designed to screen AA patients by whole exome sequencing. However, the AA is a very heterogeneous disease, and data from the entire genome has been shown to be inconclusive and very costly. Thus, we designed a comprehensive targeting sequencing panel with genes related to BMF, hematologic malignancies, DNA repair, telomere maintenance, and immune response pathways, for the screening of patients enrolled in this study (Appendix 2).

We screened two independent cohorts, one with acquired AA and another suspected to have a constitutional disease, using two different targeting sequencing approaches. The first cohort was composed of a homogenous group of 13 unrelated patients with acquired AA that underwent IST: 8 responded to treatment and 5 were refractory or clonally evolved. All of them were followed at the hematology clinic of the Clinical Hospital at Ribeirão Preto Medical School (HC-FMRP). Patients were screened for germline mutations that could modulate their IST response with our customized targeting sequencing panel. A second cohort was composed of 59 patients suspected to have a constitutional BMF followed at the National Heart, Lung, and Blood Institute, National Institutes of Health (NHLBI-NIH). Patients were screened for germline mutations in genes related to BMF with a commercial Inherited Bone Marrow Failure Sequencing panel (49 candidate genes; Appendix 3) performed by the University of Chicago Genetics Services Laboratory.

The variants identified in both cohorts were assessed for its pathogenicity and contribution to patients' phenotype. The present work is linked to the projects previously approved by FAPESP (process numbers 2014/27294-7 and BEPE 2015/19074-0).

AIMS

2. AIMS

2.1. Main aim

To screen both acquired and constitutional aplastic anemia patients using NGS approaches in order to identify genetic factors that are etiologic in patients' disease or can modulate IST response.

2.2. Specific aims

- To screen AA patients that were refractory to at least two courses of IST (n=8) and patients whose respond after one course of IST (n=5) using a customized targeting NGS panel;
- To screen AA patients from the Hematology Branch of the NHLBI-NIH (n=59) with a suspected constitutional bone marrow for bone marrow failure-related genes using targeting sequencing;
- To analyze the results obtained from the NGS using bioinformatics tools;
- To evaluate the potential impact of identified germline variants on cell signaling and cell maintenance pathways of patients with acquired or constitutional AA.

MATERIAL AND METHODS

3. MATERIAL AND METHODS

3.1. Patients and controls

We performed the genetic screening of two independent cohorts: the FMRP cohort with acquired AA and the NIH cohort with constitutional AA. In the FMRP cohort, we collected EDTA peripheral blood samples from AA patients refractory to at least two courses of IST (n=5) and responders after one course of immunosuppression (n=8). Patients with TL under the 10^{th} or 1^{st} percentile of age-matched controls or with identified genetic defect were excluded from our cohort to avoid the inclusion of patients with suspected constitutional AA. The criteria for hematologic recovery were preconized by NIH and adopted by the hematology clinic of the Clinical Hospital at Ribeirão Preto Medical School (HC-FMRP). All patients were recruited in the 'bone marrow failure clinic' at the hospital. The blood samples were collected under approval of the local research ethics committees (CAAE number 40458114.8.0000.5440, Appendix 4), following written informed consent. Blood samples from five healthy individuals were used as controls in our study. Peripheral blood samples were separated by CD3 microbeads (MACS, Germany) into myeloid fraction and CD3 positive T-cells, used for detection of 'germline' variants. The DNA was extracted from these fractions using the automated Maxwell 16 blood DNA purification kit (Promega, USA) and the DNA extraction kit (Qiagen, USA). DNA samples were quantified, stored at -20ºC, and transferred to the NHLBI-NIH for sequencing.

In the NIH cohort, we collected peripheral blood in heparin tubes (5-10 mL) from 59 patients with BMF at diagnosis routinely investigated for germline mutations and telomere defects after written informed consent in accordance with the Declaration of Helsinki and under protocols approved by the Institutional Review Board of the National

Heart, Lung, and Blood Institute. Patients were selected for the study based on a personal or familial medical history suggestive of constitutional BMF (features included early hair graying, liver cirrhosis, pulmonary fibrosis, chronic cytopenias, clonal evolution to MDS, and/or family history of BMF or hematopoietic neoplasms). Peripheral blood mononuclear cells (PBMCs) from each patient were isolated by Ficoll gradient centrifugation and separated into two aliquots. One aliquot was immediately subjected to DNA extraction with the Maxwell 16 Blood DNA Purification kit using the automated Maxwell instrument (Promega, USA) for sequencing and TL measurement, followed by storage at -20 °C. The other was stored in liquid nitrogen using freezing medium (20% fetal bovine serum [FBS] and 10% DMSO) and thawed in case of need. All DNA samples were quantified using the NanoDrop (Thermo Fisher Scientific, USA) and assessed for integrity using agarose gel electrophoresis. gDNA sequencing was carried out by the University of Chicago Genetic Services Laboratory: first, a targeted NGS panel (Appendix 3); and then Sanger sequencing for variant confirmation.

The diagnosis of AA was defined according to standard criteria (Young *et al.*, 2006; Bennett and Orazi, 2009; Killick *et al.*, 2016).

3.2.Telomere length measurement

TL was measured by Southern blot (SB) of terminal restriction fragment (TRF) or by Telomere Fluorescence In Situ Hybridization and Flow Cytometry (flow-FISH), as previously described (Gutierrez-Rodrigues *et al.*, 2014). Briefly, SB was performed using the TeloTAGGG Telomere Length Assay kit (Roche, USA), following the manufacturer's instructions. DNA (800 ng) extracted from peripheral blood cells was digested with *Hinf*I and *Rsa*I restriction enzymes. Fragments were separated on agarose gel electrophoresis and transferred to a nylon membrane. TRFs were hybridized with a digoxigenin-labeled

probe and detected with anti–DIG-AP antibody, followed by detection of the chemiluminescence on the ImageQuant LAS-4000 Analyzer (GE Healthcare, USA). A mean TRF length was analyzed using the ImageQuant TL software (GE Healthcare) by comparing signals relative to a molecular weight standard. All statistical analyses were carried out using the GraphPad PRISM version 6.0 (GraphPad Software, USA).

For flow-FISH, white blood cells (WBCs) were isolated, counted, aliquoted, and frozen at -80°C in 10% DMSO. WBCs were divided in four replicate tubes in each experiment. To control tube-to-tube variation, 10^5 fixed bovine thymocytes (CT) previously isolated were added to each sample as an internal reference. Telomere PNA Kit/FITC probe (Dako, Denmark) was used for hybridization, according to manufacturer's instructions. All samples were analyzed in a JSAN flow cytometer (BayBioscences, Japan). FITC-labeled fluorescent calibration beads (Quantum FITC-5 MESF; Bangs laboratories, Inc., USA) were used to calibrate the flow cytometer and to translate results into standard fluorescence units, as described. Using the Quantum FITC MESF software (Bangs Laboratory), the fluorescence recorded for each sample was converted into equivalent MESF value (Baerlocher *et al.*, 2006; Gutierrez-Rodrigues *et al.*, 2014). To transform MESF values into kilobases, we applied the equation described by (Kapoor and Telford, 2004). A reference sample was included as a control in each flow-FISH experiment. All the measurements were normalized by the TL calculated for the CTs added in each tube sample, as previously described (Gutierrez-Rodrigues *et al.*, 2014).

Patients' TL were compared to age-matched TL from 180 healthy individuals and defined as very short if below the first percentile and short telomeres if below the tenth percentile of distribution curves derived from controls according to age (Gutierrez-Rodrigues *et al.*, 2014).

3.3. FMRP cohort

3.3.1. Targeting sequencing

3.3.1.1. Library preparation

For library preparation, we used the DNA fraction enriched for T lymphocytes (CD3 positive). CD3+ cells are commonly used as germline control for sequencing as it maturation occurs outside the bone marrow, and then, do not share the same maturation environment as clonal myeloid cells. All DNA samples were checked for its quality (OD 260/280 ratio ranging from 1.8 to 2.0 in NanoDrop, ThermoFisher, USA) and quantified by a fluorometric method (Qubit, ThermoFisher, USA). The library preparation was performed using the customized SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library (Agilent Tecnologies, USA) designed with 165 genes related to BMF, hematologic malignancies, DNA repair, telomere maintenance, and immune response pathways (Appendix 1). For each sample to be sequenced, an individual indexed library was prepared according to the manufacturer's instructions. Briefly, 1μg of DNA was sheared using the Covaris E-series (Covaris, USA) in fragments sizing 150-200bp that were further purified using the AMPure XP magnetic beads (Beckman Coulter, USA). Then, DNA fragments ends were repaired and 3' adenylated. Fragments were ligated to pair-end adaptors and amplified with a conventional PCR step. The amplified gDNA libraries were purified and hybridized with a target-specific capture library with probes specific for the genes selected in our panel (total of probes: 22345; total size of probes: 1.4 Mbp). After hybridization, the targeted molecules were captured on streptavidin beads and indexed to specific barcodes that differs in it sequence for multiplexed sequencing (indexes A1-H12 provided by

manufacturer). The library probes were designed complementary to the *H. sapiens* genome hg19, GRCh37, February 2009. The final indexed library quality and quantity were assessed by High Sensitivity DNA assay on Agilent 2100 Bioanalyzer (Agilent Technologies, USA). For multiplexed sequencing, samples were pooled in equimolar amounts according to manufacturer's instructions.

3.3.1.2. Sequencing run

Enrichment-indexed libraries from patients' DNA samples were deep sequenced on a rapid run mode of HiSeq 2500 instrument (Illumina, USA). Sequencing run was paired-end 100 bp read length (2 x 100bp) with an 8-base pairs sample-specific index. The mean coverage was 300x for each gene.

3.3.1.3. Data analysis and variants calling

Reads were aligned to the reference sequence using Burrows-Wheeler Aligner (BWA). The quality of the data set was assessed using FastQC and metrics were generated regarding data quality over the length of the reads, percentage aligning to the genome, percentage on target and average depth. Variants were called using SAMTOOLS and GATK. The validity of called variants was assessed using a series of bioinformatic filters, which consider base, sequence, and alignment quality metrics, the percentage of reads indicating a heterozygous variant, and any directional bias in the reads indicating a variant.

The ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) was used to annotate all variants and *in silico* impact prediction was performed using SIFT and Polyphen2 software, amino acid conservation, and the Combined Annotation Dependent Depletion (CADD, http://cadd.gs.washington.edu/home). Population variant allele

frequencies were determined from the Exome Aggregation Consortium (ExAC, http://exac.broadinstitute.org), and 1000 Genomes (https://1000genomes.org). A variant was considered rare if it was novel or had a minor allele frequency of less than 0.1% in ExAC or 1000 Genomes. Variants were classified as pathogenic, likely pathogenic, of uncertain significance, likely benign, and benign according to the established American College of Medical Genetics and Genomics and Association for Molecular Pathology (ACMG) consensus criteria (Appendix 1) (Richards *et al.*, 2015).

3.3.1.4. Telomerase activity assessment for novel TERT *variants*

The impact of the telomerase variants identified in two patients with acquired AA was assessed using the TRAPeze XL Telomerase Detection Kit (Merck Milipore) according to manufacturer's guidelines. VA13 cells were plated one day before transfection and grown in Dulbecco's Modified Eagle Medium (DMEM, GIBCO) containing 10% of fetal bovine serum (FBS) and 1% of Penicillin and Streptomycin (Sigma). This cell line, derived from human-lung fibroblasts, is telomerase-negative and maintains telomeres by a recombination-based pathway called alternative lengthening of telomeres (ALT). Two micrograms of pcDNA3-WT-TERC and 1µg of pcDNA3-FLAG-TERT, pcDNA3-WT-TERT, or the mutant *TERT* R358W DNA were transfected into telomerase- deficient WI-38 VA13 cells (ATCC) at 70% confluence in each well of 6-well polystyrene dishes by using X-tremeGENE HP DNA Transfection Reagent (Roche) and OPTIMEM serum free medium (ThermoFisher Scientific), according to manufacturer's instructions. Briefly, cells were scraped off the dishes 36 h after transfection, washed 3 times in PBS, and lysed with CHAPS buffer. After incubation, lysate was submitted to centrifugation for 20 minutes at 9,600 x *g*. Supernatant was collected and protein quantification were assayed in triplicate twice for each sample according to the Pierce BCA Protein Assay (Thermo

Scientific) instructions on a Versa Max (Molecular Devices). A standard curve of 2, 1, 0.5, 0.25, 0.125, 0.0625, and 0 mg/mL BSA in CHAPS was run in parallel with samples. All standard curves were linear (R^2>0.99), and all experimental values were within the linear range. We used 300ng of protein for the fluorescent telomerase repeat amplification protocol (TRAPeze XL, Millipore). PCR was performed in a Veriti PCR System (Applied Biosystem); and fluorescence was measured in a Versa Max (Molecular Devices). Each standard, control, and unknown was run in duplicate. No negative controls, including CHAPS and RNA from the pcDNA3- FLAG-transfected group, showed more than 0.03% wild-type telomerase activity.

3.4. NIH cohort

3.4.1. Targeting sequencing

3.4.1.1. Library preparation

For the NIH cohort, patients were screened for constitutional disease with the Inherited Bone Marrow Failure Sequencing panel (Appendix 3) performed by the University of Chicago Genetic Services Laboratory (Chicago, IL, USA). Peripheral blood was collected from all patients at diagnosis. The custom enrichment design targeted the coding regions and flanking intronic regions for the 49 genes associated with BMF syndromes. Patient's DNA was enriched with the Agilent SureSelect system and sequenced using paired-end 150 bp reads with an 8 bp sample-specific index on Illumina technology.

3.4.1.2. Data analysis and variants calling

Reads were aligned to the reference sequence using Burrows-Wheeler Aligner (BWA). The quality of the data set was assessed using FastQC, and metrics were

generated regarding data quality over the length of the reads, percentage aligning to the genome, percentage on target and average depth. Variants were called using SAMTOOLS, DINDEL, and GATK. The validity of called variants was assessed using a series of bioinformatic filters, which consider base, sequence, and alignment quality metrics, the percentage of reads indicating a heterozygous variant, and any directional bias in the reads indicating a variant. Variants were annotated in regard to positions in the transcript, the coding sequence, consequence for the protein, and a collection of direct and indirect evidentiary tools including Human Gene Mutation Database (HGMD), POLYPHEN, Align GVGD, SIFT, MutationTaster, and others. The variant interpretation followed the standards and guidelines for the interpretation of sequence variants from the American College of Medical Genetics and Genomics (ACMG) (Richards *et al.*, 2015). Briefly, ACMG criteria combined the following evidences to classify variants: frequencies of variants in large populations, computational and in silico predictions, functional data, segregation analysis from family pedigrees, allelic data, function data from previous reports, and patient's phenotype. Bidirectional Sanger sequencing of mutations was performed using an ABI 3900 automated sequencer and BigDye terminator cycle sequencing reagents (Applied Biosystems) for variant confirmation.

In our hands, all potentially damaging variants were subjected to *in silico* prediction using Polyphen2 and SIFT software, analysis of 3D structure, amino acid conservation, and the Combined Annotation Dependent Depletion (CADD) tool. Population variant allele frequencies were determined from the Exome Aggregation Consortium (ExAC, http://exac.broadinstitute.org), and 1000 Genomes (https://1000genomes.org). A variant was considered rare if it was novel or had a minor allele frequency of less than 0.1% in ExAC or 1000 Genomes.

3.4.1.3. Pathogenicity assessment of RTEL1 *variants identified by NGS*

In NIH cohort, we identified a high frequency of heterozygous *RTEL1* variants. We used the *RTEL1* variants present in ExAC database as frequency controls to determine the expected frequency of heterozygous *RTEL1* variants in a large population. We filtered *RTEL1* variants present in ExAC (data from 62500 individuals) by frequency (below 0.1%) and type of variants (loss of function, missense, frameshift, and stop gained) and scored them for pathogenicity using the CADD tool. The CADD integrates different types of genomic information to score in a single value the human single-nucleotide variant (SNV) or small insertion-deletion (indel) for deleteriousness. The score defines a rank of a deleterious effect of a given variant in log (CADD of 10 and 20 assign the top 10% and 1% deleterious variants for a reference genome, respectively) (Kircher *et al.*, 2014).

3.4.2. Functional analysis

We identified novel and rare *RTEL1* variants in our cohort with constitutional AA patients. To evaluate the impact of these variants in the protein function, we performed the following functional assays.

3.4.2.1.Single-stranded 3' overhang measurement

To investigate impact of *RTEL1* variants in telomere structures, telomeric 3' overhang was assessed by non-denaturing SB as previously described (Wu *et al.*, 2012; Takai *et al.*, 2016). The protocol was adapted to use the TeloTAGGG Telomere Length Assay kit instead of the g-^{32}P-ATP end-labeled [AACCCT]4 probe for 3' overhang signals detection. 8 µg of DNA from each sample were diluted to concentration of 166 ng/µL and aliquoted as follow: (1) 18 µL (3 µg of DNA) were used for 3' overhang measurement by non-denatured SB, (2) 18 µL (3 µg of DNA) were digested with Exo1 (negative

control) and then used for 3’ overhang measurement by non-denatured SB, (3) 5 µL (800 ng of DNA) were used for total telomere detection (DNA input control) by denatured SB, (4) 5 µL (800 ng of DNA) were digested with Exo1 and then used for total telomere detection by denatured SB, and (5) 2 µL (330 ng of DNA) were run in an 0.8% agarose gel to check DNA degradation and sample loading.

For 3’ overhangs measurement, the DNA aliquot used as negative control was digested with 3 µL of Exo1 (NEB, USA) for 1 h at 37º C. Then, both Exo1 treated sample and DNA aliquot used for 3’ overhang detection was digested with FastDigest *Rsa*I/*Hinf*I restriction enzymes (ThermoFisher, USA) at 37°C for 30 min and subjected to agarose gel electrophoresis for 4 h at 90V. Samples on a gel were immediately transferred to a nylon membrane without denaturation and neutralization treatments to allow detection of 3’ overhang but not duplex telomeric strands. Telomeric probes hybridized to 3’ overhangs were detected by chemiluminescence using the ImageQuant LAS-4000 Analyzer. 3’ overhang intensity was analyzed with ImageQuant TL software. 3’ overhang signals were determined by a sum of chemiluminescent signal (ΣODi) for each sample. Background signals (sample treated with Exo1) were subtracted from total ΣODi for all the samples. Independent analyses were performed for each sample when DNA was available.

Relative 3’ overhang signals were determined by normalizing the sum of ΣODi from each individual in the non-denaturing membrane (overhang signals) by the telomeric signal in the denatured membrane (representing total genomic DNA). Telomeric signals were detected by SB of TRF, as previously described (Gutierrez-Rodrigues *et al.*, 2014). Patients’ 3’ overhang signals were normalized by an average of 3’ overhang signals from healthy individuals used as controls in every experiment. Patients’ 3’ overhang signals below the 95% confidence interval of 3’ overhang measurements from controls were

considered eroded. All samples at 166 ng/µL were run in agarose gels before each experiment and only used if DNA was not degraded and loading were similar to controls.

3.4.2.2. Lentivirus infection and culture conditions of 293T cells

The 293T cell line (ATCC) was cultured in DMEM supplemented with 10% FBS and penicillin-streptomycin-glutamine. The full-length human RTEL1 (1300 aa; NM_001283009.1) GenEZ ORF Clone (OHu14298C; GenScript) was used as an RTEL1 reference sequence (WT) and as a template for mutagenesis service provided by the company. The ORFs carrying the RTEL1 WT sequence, the P82L, M652T, D719N, G951S, and F1262L variants were cloned into the pLV-[Exp]-Puro-CMV vector for lentivirus production performed by Cyagen, Bioscience. 293T cell lines stably expressing RTEL1 were established by infecting pLV-[Exp]-Puro-CMV-RTEL1-FLAG into these cells followed by puromycin selection (3 µg/µL) 3 days after infection. Bulk 293T cells stably expressing RTEL1-FLAG (as determined by western blot) were single-cell isolated by plating the suspended cells in a 10 cm^2 petri dish after serial dilutions. Colonies derived from single cells were carefully tripsinized after colony delimitation in the petri dish with grease and expanded to be use in further experiments. Infected 293T cells were cultured in DMEM supplemented with 10% FBS, penicillin-streptomycin-glutamine and puromycin.

3.4.2.3. Western blot and immunoprecipitation

Infected 293T cells were synchronized by the double-thymidine–block method as previously described (Sarek *et al.,* 2015). Cells were washed twice with PBS, scrapped from a 75-cm^2 flask, and collected in RIPA buffer (ThermoFisher) with EDTA-free Complete Protease Inhibitor Cocktail (Roche). Cells were lysed for 20 min on ice and

soluble whole protein extracts were obtained after centrifugation at 14000 rpm at 4º C for 15 min. Immunoprecipitation was performed using the Dynabeads™ Protein G Immunoprecipitation kit (ThermoFisher), according to the manufacturer's instruction. Whole-cell extracts (6 mg) were incubated overnight at 4º C with Dynabeads® Protein G coated with monoclonal rabbit anti-FLAG antibody (8H8L17; ThermoFisher). The FLAG antibody was crosslinked to the Dynabeads with 5 nM BS^3 (ThermoFisher) before the immunoprecipitation. The preclearing of the protein extracts is not required using this kit. Dnyabeads-FLAG-proteins complexes were washed before target elution. Immunocomplexes and whole protein extracts (50 µg) were subjected to SDS-PAGE using a 10% Tris-Glycine gel and immunoblotted to a PVDF blotting membrane using the iBind western system (ThermoFisher). Proteins were detected using the following primary antibodies: monoclonal mouse Anti-TRF2 (4A794, NB 100-56506; Novus Biologicals); monoclonal mouse Anti-FLAG M2 HRP-conjugated (A8592; Sigma-Aldrich), and monoclonal mouse beta-actin HRP-conjugated (NB600-501-H; Novus Biologicals). Peroxidase conjugated goat anti-mouse (H + L) secondary antibody (#32430; Thermo Scientific) was used to detect the anti-TRF2 antibody.

3.4.2.4. RNA extraction and PCR array

Expression levels of DNA damage signaling and telomere and telomerase related genes were assessed in patients with *RTEL1* variants from the NIH cohort. Total RNA from patients and healthy controls was extracted from frozen PBMCs using the RNeasy Plus Mini kit (Qiagen). Samples were quantified and converted to cDNA with the RT2 First Strand kit (Qiagen). Transcriptional profiles were assessed with the RT^2 Profiler PCR Array Human DNA Damage Signaling Pathway (PAHS-029ZA, Qiagen) and Telomere and Telomerase (PAHS-010Z, Qiagen), according to manufacturer's instructions. Data

analysis was performed in the RT^2 Profiler PCR Array Data Analysis version 3.5 (Qiagen). Relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. Hierarchical clustering was applied to the differentially expressed genes (fold change ≥ 2), with average linkage clustering as a linkage method. *RTEL1* expression was validated by real time PCR using the Taqman® probe for *RTEL1* (Hs01566915_m1) and a customized Taqman® probe for detection of the *RTEL1* RING domain specific for isoform 3.

3.4.2.5.*Telomere circle (TC) assay*

TC assay was performed as described previously (Zellinger *et al.*, 2007; Henson *et al.*, 2009; Sarek *et al.*, 2015), following minor modifications. Total DNA was extracted from frozen ACK-lysed peripheral nucleated blood cells by isopropanol precipitation using the Gentra® Puregene kit (Qiagen). First, samples were digested with 70U of Exonuclease V (RecBCD, NEB) for 1 h at 37º C. Digested DNA was purified by ethanol precipitation after the pH adjustment with 3 M sodium acetate. ExoV treated samples were then digested with FastDigest *Rsa*I/*Hinf*I restriction enzymes (ThermoFisher) at 37° C for 30 min, followed by enzyme inactivation at 65° C for 5 min. TC assay was performed in a reaction without cytosine (dCTP) that uses Phi29 DNA (ThermoFisher Scientific) polymerase to self-primer t-circles for rolling amplification. ExoV/*Rsa1*/*Hinf1* digested samples were incubated with 7.5U Phi29 DNA in a reaction with 1x Phi 29 Buffer and a mix of 10 mM dATP, dGTP, and dTTP at 37° C for 16 h. After enzyme inactivation at 65° C for 20 min, samples were diluted (1:2) with water. Half of the diluted Phi29 product were mixed with 5x Loading buffer and subjected to a 0.8% agarose gel for SB analysis. T-circles were detected using the TeloTAGGG Telomere Length Assay kit and represented by a high molecular weight C-rich telomeric product, since both genomic and G-rich telomeric fragments are not amplified in a reaction without dCTP. A

negative control represented by a Phi29 negative sample was included in the assay. A chemiluminescence signal from the digoxigenin-labeled telomeric probe labeled with an anti–DIG-AP antibody was detected using the ImageQuant LAS-4000 Analyzer (GE Healthcare).

RESULTS

4. RESULTS

4.1. FMRP cohort

We enrolled for this study 13 unrelated patients (NGS_01 to NGS_13; median age of 49 years old) with severe AA that underwent IST therapy. Eight patients presented a complete response to IST after 1 year of treatment, four were refractory to at 2 cycles of IST, and one showed partial response but clonally evolved to monosomy 7 after treatment (Table 1). Patient NGS_07 was refractory to IST and deceased after evolution to AML. We also included in our study five healthy individuals that were used as controls.

IST regimens were different among patients based on treatment available at the time of diagnosis. Five patients were treated with rabbit ATG and other five with a horse ATG, both concomitant with ciclosporin. Patients' TL was normal, except for patient NGS_10 that presented TL in the borderline of the 10^{th} percentile of age-matched controls.

Patients' peripheral blood samples separated by CD3 microbeads were quantified and assessed for it quality prior sequencing (Table 2). All DNA samples were within accepted range of quality and were then used in library preparation step.

As expected, hematologic counts and TL were normal for all the controls. The DNA extracted from both CD3 positive and negative fractions presented high quality and concentration (Table 2). For sequencing, we selected two control samples (CT4 and CT6) that were processed and sequenced in parallel with the patients' samples.

Table 1. Clinical features of patients with acquired aplastic anemia from FMRP cohort enrolled in the study.

ID	Age	Diagnosis	IST regimen	IST response	Hematologic counts after IST		
					Hb (g/dl)	Neutrophils ($x10^3$ mm^3/µL)	Platelets ($x10^3$ mm^3/µL)
NGS_01	23	SAA	rATG+ CSA	Complete	13.9	1.8	77
NGS_02	35	SAA	rATG+ CSA	Complete	12	2.1	70
NGS_03	68	SAA	rATG+ CSA	Refractory†	8.2	0.1	1
NGS_04	18	SAA	hATG+CSA	Clonally evolved to monossomy 7	7.7	1.2	31
NGS_05	37	SAA	hATG+CSA	Complete	13.8	2.1	218
NGS_06	49	SAA	hATG+CSA	Complete	16.7	1.6	184
NGS_07	77	SAA	rATG + CSA/eltrombopag	Refractory/ Evolved to AML†	7.2	0.4	20
NGS_08	61	SAA	hATG+CSA	Complete	14.1	3.1	272
NGS_09	60	SAA	hATG+CSA	Complete	15.2	2.6	149
NGS_10	51	SAA	rATG+CSA	Refractory	6	0.5	8
NGS_11	37	SAA	rATG+CSA	Complete	13.2	2.8	224
NGS_12	54	SAA	rATG+CSA	Refractory†	7.5	0.2	18
NGS_13	50	SAA	rATG+CSA	Complete	N.A	N.A	N.A

All samples were collected at least 1 year after IST. The criteria for complete response were platelets >50000/mL, hemoglobin (Hb) >9g/dL, and neutrophils >800 mm^3/µl. SAA, severe aplastic anemia; IST, immunosuppressive therapy; rATG, rabbit antithymocyte globulin; hATG, horse antithymocyte globulin; CSA, cyclosporine; †Deceased.

Table 2. Samples' quality for targeting sequencing.

ID	Age	IST response	DNA concentration (ng/µl) [260/230; 260/280 ratios]				
			Extraction by Maxuell ®		Qiagen ® Puregene manual protocol		
			Leukocytes	Myeloid fraction	Leukocytes	T cell CD3+ fraction	Myeloid fraction
NGS_01	23	Complete	50.0 (1.8/1.7)	N.R	N.R	N.R	N.R
NGS_02	35	Complete	32 [2.0/1.8]	N.R	N.R	N.R	N.R
NGS_03	68	Refractory	31 [1.9/1.8]	N.R	N.R	N.R	N.R
NGS_04	18	Clonally evolved to monosomy 7	116.6 [2.1/1.9]	344.16 [1.4/1.8]	N.R	342.1 [1.8/1.8]	342.1 [1.8/1.8]
NGS_05	37	Complete	33 [2.0/1.8]	273.46 [1.8/1.8]	N.R	267.4 [1.9/1.7]	267.4 [1.9/1.7]
NGS_06	49	Complete	13 [1.8/2.0]	333.0 [1.8/1.9]	169.1 [2.0/1.9]	80.32 [2.0/1.9]	80.32 [2.0/1.9]
NGS_07	77	Refractory/ Evolved to AML	N.R	74.6 [2.3/1.9]	N.R	162.4 [2.2/1.9]	162.4 [2.2/1.9]
NGS_08	61	Complete	80 [2.3/1.9]	259 [0.8/1.8]	523.8 [2.1/1.8]	573 [2.1/1.8]	573 [2.1/1.8]
NGS_09	60	Complete	108 [2.2/1.9]	N.R	N.R	82.9 [2.0/1.9]	524.9 [2.2/1.8]
NGS_10	51	Refractory	73 [2,2/1,9]	N.R	N.R	106,5 [2,2/1,8]	516,2 [2,2/2,8]
NGS_11	37	Complete	204.6 [1.9/1.8]	394.08 [1.1/1.9]	N.R	66 [2.1/1.96]	457 [2.2/1.88]
NGS_12	54	Refractory	N.R	N.R	N.R	61.8 [1.9/1.9]	473.5 [2.1/1.88]
NGS_13	50	Complete	72.2 [1.8/1.9]	N.R	N.R	N.R	N.R
CT2	33	Normal control	244 [1.5/1.8]	177.89 [1.66/1.8]	N.R	N.R	N.R
CT3	53	Normal control	164.79 [2.3/1.8]	208.33 [1.6/1.8]	N.R	N.R	N.R
CT4	26	Normal control	183.3 [0.83/1.8]	261 [1.5/1.8]	N.R	N.R	N.R
CT5	26	Normal control	81.57 [1.85/2.1]	138.4 [1.35/1.84]	N.R	N.R	N.R
CT6	24	Normal control	113.2 [1.82/1.88]	N.R	N.R	N.R	N.R

Peripheral blood samples were separated by CD3 microbeads (MACS. Miltenyi. Germany) into myeloid cells used as "tumor" cells and CD3 positive T-cells, used as "germline" sample. All DNA samples were store at -20°C. N.R= not performed; N.A= not available. The DNA from fractioned cells was extracted using two different methodologies based on the number of cells recuperated after magnetic separation: an automated extraction using the Maxwell instrument (Promega, Madison, USA) or a manual extraction using the Gentra Puregene Blood kit (Qiagen, Maryland, USA).

4.1.1. Library preparation

We designed a customized targeting sequencing panel with 165 genes related to BMF, hematologic malignancies, DNA repair, telomere maintenance, and immune response pathways (Appendix 2). In the first step of the protocol, 1µg of CD3-enriched DNA from 13 AA patients and two healthy controls were sheared in Covaris, purified using the Ampure XP magnetic beads, and assessed for its quality in Bioanalyzer. After shearing, the average size of fragments from all samples was 212pb (Figure 4A). The mean recuperation rate after magnetic purification was 45% of the initial DNA input. DNA fragments were then repaired at the ends, 3' adenylated, ligated to paired-end adaptors before hybridization with customized probes. The complex DNA-probe were captured and indexed for sequencing. The average size of indexed libraries processed was 358bp (Figure 4B) and final concentration ranged from 13nM to 33nM assessed by 2100 Bioanalyzer. For sequencing, 15 indexed libraries were pooled in equimolar amounts in a final volume of 100 µL at 20 nM (concentration measured by Qubit = 8.4 ng/µL).

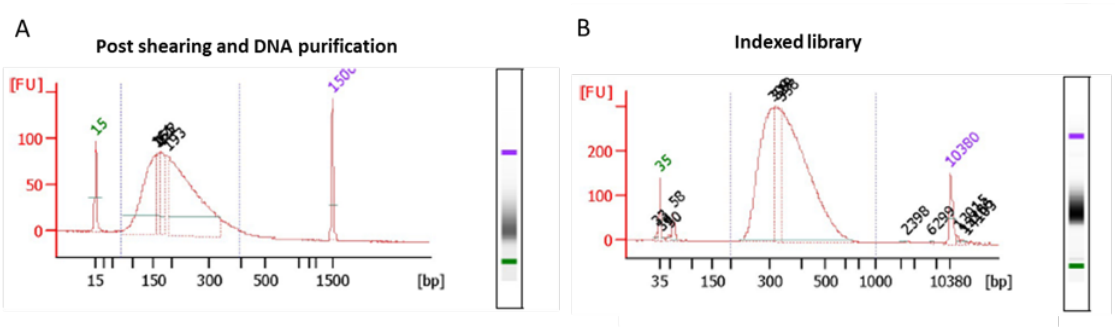


Figure 4. Quality analysis of DNA samples in library preparation protocol using the 2100 Bioanalyzer. (A) Representative analysis of sheared DNA using a DNA 1000 Bioanalyzer assay. Mean fragment size in this sample was 208 bp, ranging from 73 to 394 bp. **(B)** Representative analysis of post-capture indexed library DNA using the High Sensitivity assay. Mean fragment size in this sample was 366 bp, ranging from 200 to 668 bp.

4.1.2. Targeting sequencing and data analysis

All indexed libraries had a high phred quality score (Q > 30) with reads lengthening 100 bp and GC% content around 40-42. The data (FASTQ files) were aligned to reference genome (H. sapiens genome hg19, GRCh37) and then recalibrated using GATK tools to generate the BAM file ready to analysis. Variants were called using HaplotypeCaller and filtered based on variant quality score recalibration (VQSR) from GATK package tools. Filtering was refined according to allele frequency in population databases (allele frequency >0.1% in ExAC and 1000 genomes) and The Cancer Genome Atlas Network database (filtered TCGA variants). Using ANNOVAR, the annotation of genetic variants was based on the canonical transcript and examined their functional consequence on genes by amino acid conservation analysis, in silico deleteriousness prediction, or identification in single nucleotide polymorphisms (SNPs) databases.

We annotated 3906 rare variants in our cohort, including missense, intronic, 5' and 3' – UTR, upstream and downstream, and splicing variants. Further analysis revealed that only missense variants had potential to be disease-causing, and thus, these variants were reported for each patient (Table 3). We screened GATA2 intronic regions for mutations (Hsu *et al.*, 2013), but we found no variants in our cohort. No functional relevant variants were identified in our controls

Table 3. Mutational profile of aplastic anemia patients.

Patient	Sex	Age	Gene	Nucleotide change	Amino acid change	ACMG classification	ENSEMBL	ExAC Browser frequency	In silico analysis*	CADD PHRED score
Responders										
NGS_01	Male	23	*FANCI*	c.2131 T>G	V682G	Uncertain Significance	ENST00000310775	NA	Possibly benign	13.5
NGS_02	Female	35	*RTEL1*	c.1287 C>T	A290V	Uncertain Significance	ENST00000370018	<0.001	Possibly benign	9.3
NGS_05	Female	37	*WAS*	c.1187 C>T	P371L	Uncertain Significance	ENST00000376701	NA	Indeterminate	22.2
NGS_06	Male	49	*PEG3*	c.2386 G>T	R674S		ENST00000326441	NA	Probably damaging	22.6
			NCOR2	c.4055 C>T	R1352C	Uncertain Significance	ENST00000405201	NA	Probably damaging	34
			PML	c.1759 C>T	R555C		ENST00000268058	NA	Possibly benign	18.4
NGS_08	Female	61	No variants identified							
NGS_09	Male	60	No variants identified							
NGS_11	Female	37	No variants identified							
NGS_13	Female	50	*FLT3*	c.2658 G>A	S859N	Uncertain Significance	ENST00000241453	NA	Probably damaging	29.2
Nonresponders / Clonally evolved										
NGS_03	Female	68	No variants identified							
NGS_04	Male	18	*MLL*	c.11283 G>A	E3754K	Uncertain Significance	ENST00000534358	NA	Probably damaging	34
			DHX36	c.1641 A>C	L520F		ENST00000496811	NA	Probably damaging	23.4
NGS_07	Male	77	*TERT*	c.1072 C>T	R358W	Likely pathogenic	ENST00000310581	NA	Probably damaging	24.4
NGS_10	Male	51	*CALR*	c.95 G>C	L8V		ENST00000316448	NA	Indeterminate	11.4
			SETBP1	c.2508 C>T	P738S	Uncertain Significance	EST00000282030	NA	Possibly benign	9.2
			GNAS	c.964 C>T	L138F		ENST00000371100	NA	Indeterminate	0.001
NGS_12	Male	54	*TERT*	c.1072 C>T	R358W	Likely pathogenic	ENST00000310581	NA	Probably damaging	24.4
			GNAS	c.1682 G>T	G377V	Uncertain Significance	ENST00000371100	<0.0001	Indeterminate	0.001

4.1.3. Mutational profile of patients with acquired aplastic anemia

Five out of 8 patients that responded to IST carried germline variants in genes screened in our panel (Table 3), while in three patients no potentially damaging variants were identified. Except for the *RTEL1* variant in NGS_2, variants were not previously reported or stored in genome databases. In this group, patients NGS_6 and NGS_13 carried variants predicted as damaging in silico. So far, only somatic variants in *PEG3*, *NCOR2*, and *FLT3* genes have been associated with hematologic disorders. Somatic variants in *PEG3* were identified in a Japanese cohort with AA (Yoshizato *et al.*, 2015) and both *NCOR2* and *FLT3* are recurrent mutated in AML/MDS but not common in AA. Dysfunctional *NCOR2* gene appears to be more related to impaired hematopoiesis rather than malignant diseases. In zebrafish, *NCOR2* is required for hematopoietic stem cell development (Wei *et al.*, 2014), while in hematologic malignancies this gene appears to play a crucial role in leukemogenesis (Hong *et al.*, 2001). Somatic acquisition of internal tandem duplications and point mutation in codon 835 of *FLT3* are very frequent in AML; however, in our study, the potentially deleterious impact of FLT3-S859N may not be associated with loss of function, but instead with constitutive activation of the receptor. Nevertheless, the clinical significance of these variants in AA remains unknown.

In contrast, three out of 5 patients that were refractory to IST, clonally evolved to AML/MDS or developed monosomy 7 carried germline variants in genes related to telomere biology or myeloid neoplasia (Table 3). Only NGS_3 was refractory to IST and not identified with any functionally relevant germline variants. The variants identified in NGS_10 were not clinically related to patient's phenotype, even so, the *SETBP1* are find somatically mutated in AML/MDS. NGS_10 was the only patient with TL in 10th

percentile boundary, but his accelerated telomere loss may likely secondary to proliferative stress.

A novel *TERT* R358W (c. 1072 C>T) classified as likely pathogenic was identified in two unrelated patients (NGS_07 and NGS_10) (Table 3). In both cases, these variants were interpreted as a possible mosaic, since the recalibrated BAM files displayed these variants with frequencies of 20% and 32% in NGS_07 and NGS_10, respectively.

Surprisingly, this novel variant was already identified by our group in one family with telomeropathy followed at FMRP/USP (Figure 5). The index case, an 8 years old female patient, was homozygous for this mutation and presented a severe AA and very short telomeres. She also harbored a homozygous *TERT* A279T (c.835 G>A) and G677G (c.2031 C>T) polymorphisms.

We assessed the impact of this mutation on telomerase function by measuring the telomerase activity in VA13 transfected with a plasmid mutated for R358W, A279T, and G677G variants and with the R358W mutation alone. Cell lysates with R358W variant presented 72% of the wild-type activity (Figure 6). No significant differences ($p>0.05$) were seen with telomerase activity between both constructs, evidencing the telomerase impairment was due to the R358W mutation (Figure 6). The proband's father, mother, and sister were heterozygous for R358W, A279T, and G677G variants (Figure 5A). However, they had normal telomeres and were asymptomatic, which could be explained by the fact of being heterozygous for a mutation that maintains more than 50% of telomerase activity. For R358W, we had assumed that both alleles must be compromised to develop the clinical manifestations of AA, and thus may be associated with an autosomal recessive form of the disease.

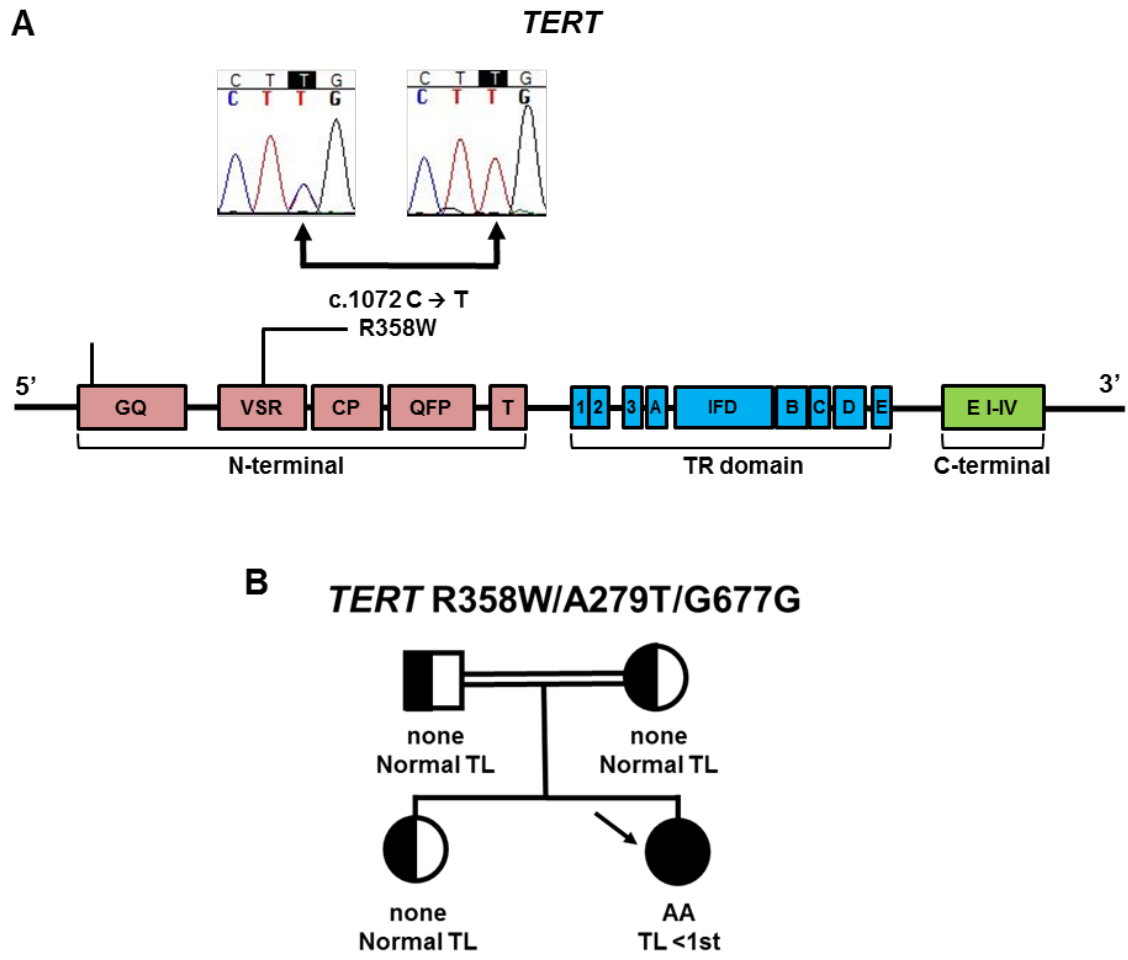


Figure 5. A novel *TERT* R358W mutation. (A) Linear representation of *TERT* gene. This gene contains different domains and is divided into the N-terminal, C-terminal and transcriptase reverse regions. The R358W variant, identified in either homozygosis or heterozygosis, was confirmed by Sanger sequencing and are represented by chromatograms. **(B) Pedigrees of the family identified with the *TERT* variant.** The *TERT* variant present in the family is described above pedigree. Arrows indicated the proband. Half or full black circles and squares represent individuals who are heterozygous or homozygous for the *TERT* variant. When tested, telomere length (TL) and clinical features are described in the figure under each individual.

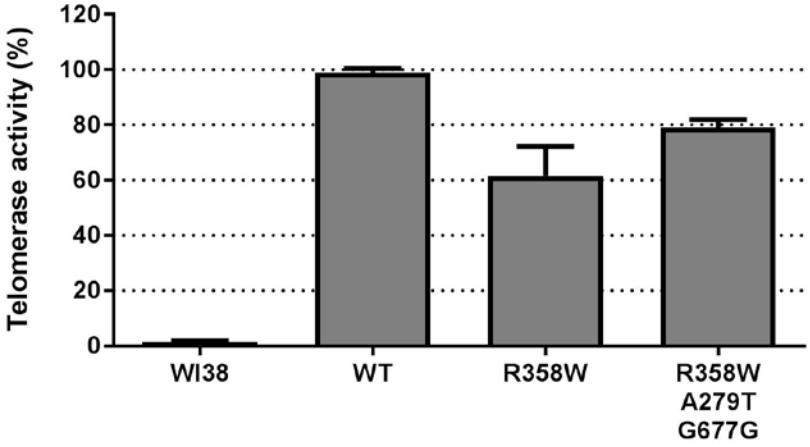


Figure 6. The telomerase activity of the novel *TERT* R358W mutation. The impact of the telomerase mutations was assessed with cell lysates obtained from WI38-VA13 cells transfected with the mutated and wild-type vectors using the TRAPeze XL kit (Millipore, USA). Telomerase activity was considered 100% for the wild-type. Each bar for each vector represents the mean of three independent experiments with the standard deviation for all experiments for each vector. No differences were seen in telomerase activity from the R358W vector and the R358W associated with two polymorphisms (G677G and A279T) also present in the proband ($p>0.05$).

The minimal reduction in telomerase activity caused by a somatic R358W clone may impair the HSC capacity to repopulate the bone marrow in stress conditions, as seen after IST. In that case, it is possible that R358W can modulate hematopoiesis and response after IST. NGS_7 was refractory to IST but responded to eltrombopag after 6 months of treatment. Later, this patient evolved to AML and deceased. Although 95% of patients treated with eltrombopag respond after 6 months of treatment (Townsley *et al.*, 2017), it is unknown if patients with germline defects are prone to clonally evolved during follow up.

Patient NGS_4 evolved with monosomy 7 after two courses of IST and carried variants in the genes *MLL* and *DHX36*; both predicted pathogenic and with very high CADD score (Table 3). *MLL* regulates gene expression during hematopoiesis and is a recurrent target in myeloid leukemia, mostly associated with somatic rearrangements that

lead to the production of MLL associated oncogenic fusion proteins (Rao and Dou, 2015). MLL haploinsufficiency in mice leads to developmental delay and reduced blood counts, but in humans, heterozygous *MLL* variants are associated with Wiedemann-Steiner Syndrome, in which patients have no predisposition to cancer (Jones *et al.*, 2012). In AA patients, oligoclonal hematopoiesis and somatic acquisition of mutations in AML/MDS driver genes precede monosomy 7 (Dumitriu *et al.*, 2015). However, in our study, the role of *MLL* E3754K variant in patient's phenotype needs to be assessed.

NGS_4 also harbored a novel variant L520F in *DHX36* gene. This gene encodes a DEAH-box family RNA helicase (DDX36) that was recently found to play a role in telomere maintenance (Lattmann *et al.*, 2011; Sexton and Collins, 2011; Booy *et al.*, 2012). This RNA helicase binds to the G4 quadruplexes motif at the 5' region of *TERC* to promote RNA accumulation through unwinding the 5' G4 structures in TERC and formation of the P1 helix template boundary critical for telomerase fidelity (Sexton and Collins, 2011; Booy *et al.*, 2012). DDX36 also binds to TERT through interaction with TERC, which gives access to a resolution of G4 structures at telomeric DNA. The interaction between DDX36 and G4-DNA at telomeres requires a free 3' guanine-containing tail, such as the 3'overhangs, that provides access to telomere replication by telomerase (Smaldino *et al.*, 2015). In our study, the L520F is predicted as damaging in silico, have high CADD score and are located within the Helicase-C domain on the enzyme, all evidence of the pathogenic effect of this variant in DDX36 function.

Despite being associated with normal TL, this novel *DHX36* variant may impair patient's telomere maintenance and modulate his response to treatment. Also, this helicase stood as a new component critical to telomere biology and related to telomere diseases.

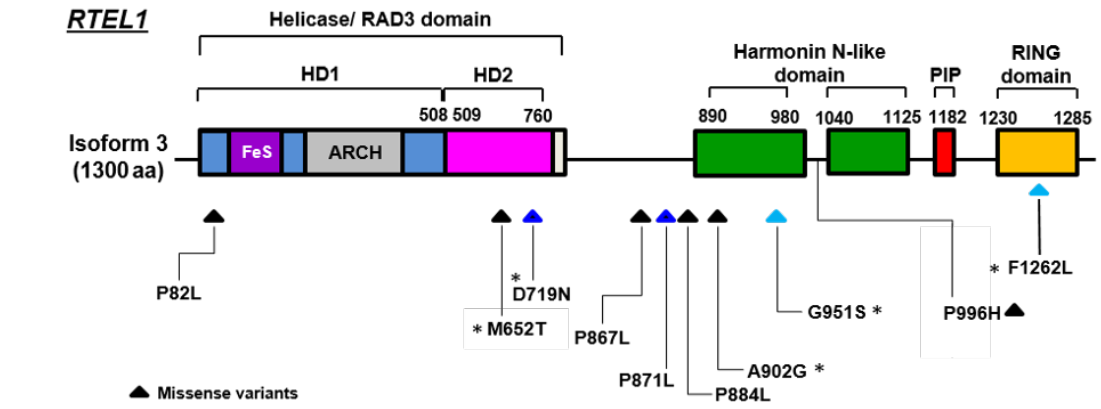
4.2. NIH cohort

4.2.1. Mutational profile of patients with constitutional aplastic anemia

We analyzed mutational data from 59 patients with BMF followed at NIH that were consecutively screened for germline mutations in genes related to BMF as well for TL. Using this approach, patients with both acquired or constitutional BMF have been screened in parallel.

We notice a high frequency of patients carrying heterozygous *RTEL1* variants in this cohort. We identified 10 germline *RTEL1* variants among 8 patients, including 2 with biallelic variants (Figure 7). The *RTEL1* encodes a helicase critical to genome integrity, DNA repair, and telomere maintenance: it disassembles the t-loops and G4 quadruplexes (DNA secondary structures) at telomeres required for replication (Vannier *et al.*, 2012). Full hematologic data and clinical features are shown in Table 4 and 5.

A



B

Potentially deleterious *RTEL1* variants # alternative alleles (percentage of population)		
	ExAC*	NIH
Missense	1248 (2.05)	10 (16.9)
LOF	25 (0.04)	0 (0)

* *RTEL1* variants with allele frequency in ExAC or 1000 Genomes less than 0.1%

C

Stratification of rare missense *RTEL1* variants in ExAC and NIH cohort by CADD phred score # alternative alleles (percentage of population)		
	ExAC	NIH
CADD under 15	728 (1.19)	7 (11.8)
CADD over 15	520 (0.85)	3 (5.08)
CADD over 20	389 (0.64)	3 (5.08)
CADD over 30	60 (0.09)	0 (0)

Rare *RTEL1* variants: novel or allele frequency less than 0.1%.

Figure 7. Linear representation of RTEL1 isoforms 3 (1300 amino acids; NCBI NM_016434.3). (A) Colored boxes indicate the conserved RTEL1 domains and lines indicate the position of each variant in the gene. Ten novel or rare heterozygous *RTEL1* variants were identified in 8 patients from NIH cohort (n = 59). The number of patients carrying each variant are represented by triangles, which also represents the type of the variant (missense variants). Blue triangles represent 2 compound heterozygous patients that carried 2 different *RTEL1* variants. Five variants depicted with an asterisk (*) had CADD phred score > 15 (this score was selected to predicted RTEL1 deleteriousness). The heterogeneity of RTEL1 functions is attributed to its different domains: RAD3, the catalytic core responsible for DNA unwinding activity; PIP, required for binding to proliferating cell nuclear antigen (PCNA) and promotion of G4 quadruplexes unwinding; C4C4-RING, which resolves t-loops during telomere replication; and the Harmonin N-like, a region of unclear function (Vannier *et al.*, 2014). **(B)** Frequency of potential deleterious *RTEL1* missense and loss of function (LoF) in ExAC database (n = 60706), and NIH (n = 59) cohort. Missense variants are enriched in the NIH cohort as compared to ExAC database. **(C)** Frequency of missense *RTEL1* variants stratified by CADD phred score in ExAC database and NIH cohorts. The CADD predicts deleteriousness of a given variant. All variants with CADD over than 15 NIH had CADD over than 20 as well. The CADD of 15 was selected as a cut off to predict a variant as deleterious based on the frequency identified in ExAC database.

Table 4. Description of 27 patients with *RTEL1* variants

Patient ID	Age/ Sex	TL	*RTEL1* variants	Other germline variants	*RTEL1* ACGM *	Diagnosis	Other clinical features	Family history	Outcome
Heterozygous pathogenic or likely pathogenic variants									
NIH-4	31/F	Normal	M652T, c.1955 T>C	*SLX4*, T750M c.2249 C>T	LP	ICUS - neutropenia	Normal BM cellularity	None	No treatment
Biallelic pathogenic or likely pathogenic variants									
NIH-1	32/M	<1st	D719N, c.2155 G>A P871L, c.2612 C>T	None	LP LB	MAA	Liver cirrhosis, pulmonary fibrosis	Pedigree (Figure 9)	Responded to androgen (danazol)
NIH-2	6/F	<1st	G975S, c.2923 G>A F1262L, c.3786 C>G	None	LP P	SAA	Prolonged thrombocyto-penia	Pedigree (Figure 9)	Awaiting HSCT
Variants of uncertain significance									
NIH-3	17/M	Normal	P884L, c.2651 C>T	None	US	SAA	None	FH of cancer	Treated with IST/EPAG and evolved to -7/MDS
NIH-5	63/F	Normal	P82L, c.245 C>T	None	US	MAA	None	Pedigree (Figure 9)	Progressed to SAA on EPAG and evolved to MDS/AML, died following HSCT of relapsed AML

NIH-7	32/F	<1st	A902G, c.2705 C>G	†*TERC*, r.287 C>G	US	MAA	Frequent miscarriages, early hair graying	Maternal early hair graying, MDS in maternal uncle, BMF in paternal grandmother	No treatment
Likely benign variants									
NIH-6	23/F	Normal	P867L, c.2672 C>T	*TERT*, R537C [c.1609 C>T]	LB	MAA	Eczema	Father and brother with early hair graying, mother with macrocytic anemia	No treatment
NIH-8	47/M	Normal	P996H, c.2987 C>A	None	LB	hypoMDS	None	Grandfather with polycythemia vera	Responded to EPAG

The *RTEL1* variants identified in our study were annotated using the isoform 3 (1300 aa; NM_001283009.1). Abbreviations: TL, telomere length; <1st, TL below the first percentile of age matched controls (very short telomeres); <10th, TL below tenth percentile (short telomeres); BM, bone marrow; FH, family history; AML, acute myeloid leukemia; ICUS, idiopathic cytopenia of US; MAA, moderate aplastic anemia (AA); SAA, severe AA; MDS, myelodysplastic syndrome; hypoMDS, hypoplastic MDS; SLE; Systemic Lupus Erythematosus; EPAG, eltrombopag; CSA, cyclosporine; ATG, antithymocyte globulin; IST, immunosuppressive therapy with ATG and CSA; HSCT, hematopoietic stem cell transplantation; *American College of Medical Genetics and Genomics and Association for Molecular Pathology (ACMG) consensus criteria. †Variants previously reported as pathogenic.

Table 5. Hematological and molecular features of patients with *RTEL1* variants.

Patient ID	Sex	Age	WBC ($x10^3$/mL)	Neu ($x10^3$/mL)	Hb (g/dL)	MCV (fl)	Platelets ($x10^6$/mL)	Ret ($x10^9$/L)	PNH % gran)	BM cellularity (%)	BM cytogenetics
NIH-1	Male	32	1.47	0.86	12.4	108.7	21	1.7	N	10-20	46, XY [20]
NIH-2	Female	6	2.3	0.73	8.2	103.5	18	3.0	N	< 5	46, XX [20]
NIH-3	Male	17	1.43	0	10.7	80.4	33	0.3	N	< 10	46, XY [20] **
NIH-4	Female	31	1.99	0.78	12.5	95	244	2.2	N	60-70	46, XX [20]
NIH-5	Female	64	2.63	1.12	9.7	103.2	24	1.5	N	< 10	46, XX [20]
NIH-6	Female	23	3.34	1.06	10.6	111.9	83	1.9	7.5	10	46, XX [20]
NIH-7	Female	10	3.65	1.63	8.6	84.7	83	1.8	N	N. A	46, XX [20]
NIH-8	Male	47	2.52	1.18	7.1	111.2	32	4.0	11	15	46, XY [20]

WBC, white blood cells; Neu, neutrophils; Ret, reticulocytes; BM, bone marrow, Hb, hemoglobin; MCV, mean red cell volume; PNH gran, PNH granulocyte clone size; NA, not available; ** Patient evolved to monosomy 7 after initial diagnosis.

According to ACMG criteria, 4 of these 10 variants were pathogenic or likely pathogenic (M652T, D719N, G951S, F1262L), and were identified in 3 unrelated patients. Pathogenic/likely pathogenic variants were identified in 3 (5%) of 59 patients from the NIH cohort (Table 4). Three variants were of uncertain significance, as data were insufficient to fulfill ACMG criteria. Three variants classified as likely benign were identified in 2 unrelated patients and in the compound heterozygous NIH-1 that also carried a likely pathogenic variant (Table 4 and Figure 7).

Eleven variants had CADD score > 15 (this score was selected to predict RTEL1 deleteriousness based on the accepted control population frequency) and were predicted as damaging to RTEL1 functions in silico (Table 6). Three of 10 variants were absent from population databases and 6 of 10 were located in highly conserved amino acid positions (more than 80% of conservation; Figure 8).

To ascertain whether likely pathogenic *RTEL1* variants were enriched in constitutional BMF syndromes, we compared frequencies of these variants from our cohort to the ExAC database. The frequency of rare *RTEL1* missense variants in NIH cohort was increased as compared to ExAC (16.9% vs 2%, [odds ratio, 7.90 95% CI, 3.845 to 15.55]) (Figure 7B). However, not all missense variants were predicted to be deleterious. Since it is not feasible to assess the pathogenicity of *RTEL1* variants in ExAC based on ACMG criteria, we stratified *RTEL1* variants from this database based on their CADD phred scores (Figure 7C). Missense variants with CADD score higher than 20, but not >30, were still enriched in the NIH cohort as compared to ExAC (Figure 7C).

Table 6. Germline *RTEL1* variants

Patient ID	Ethnicity	*RTEL1* variant	Amino acid change Isoform 3	Amino acid change Isoform 2	ExAC Browser frequency (%)	ExAC frequency based on ethnicity (%)	Polyphen prediction	SIFT prediction	CADD score
NIH-1	White	c.2125 G>A	D719N	D743N	9/119626 (0.0007%)	3/65314 (0.004%)[1]	Benign	Tolerated	23.3
		c.2612 C>T	P871L (rs144002969)	P895L	15/119722 (0.01%)	11/65368 (0.01%)[1]	Benign	Tolerated	0.913
NIH-2	White	c.2851 G>A	G951S	G975S	0/121412	NA	Probably damaging	Deleterious	27.4
		c.3786 C>G	F1262L	x	0/121412	NA	Benign	Deleterious	23.2
NIH-3	African American	c.2651 C>T	P884L (rs144002969)	P908L	17/119146 (0.01%)	0/10204[2]	Benign	Tolerated	0.018
NIH-4	African American	c.1955 T>C	M652T (rs148080505)	M676T	17/112416 (0.01%)	0/9534[2]	Benign	Tolerated	13.46
NIH-5	White	c.245 C>T	P82L	P82L	31/119576 (0.02%)	29/65706 (0.04%)[1]	Possibly benign	Tolerated	0.03
NIH-6	White	c.2600 C>T	P867L (rs139083375)	P891L	52/119780 (0.04%)	42/65398 (0.06%)[1]	Possibly benign	Tolerated	3.8
NIH-7	White	c.2705 C>G	A902G	A926G	0/121412	NA	Benign	Tolerated	12.2
NIH-8	South East Asian	c.2987 C>A	P996H	P1020H	17/118850 (0.01%)	16/8580 (0.1%)[3]	Possibly benign	Tolerated	8.7

All patients carried the variants in heterozygosity. HD1 and HD2, helicase domain 1 and 2, respectively. ExAC frequency in 1European (Non-Finnish), 2Africans, 3East Asian population, or 4Other. CADD phred score above 15 was used to predict a variant as pathogenic. LoF, loss of function. NA, not available. x Present only in Isoform 3.

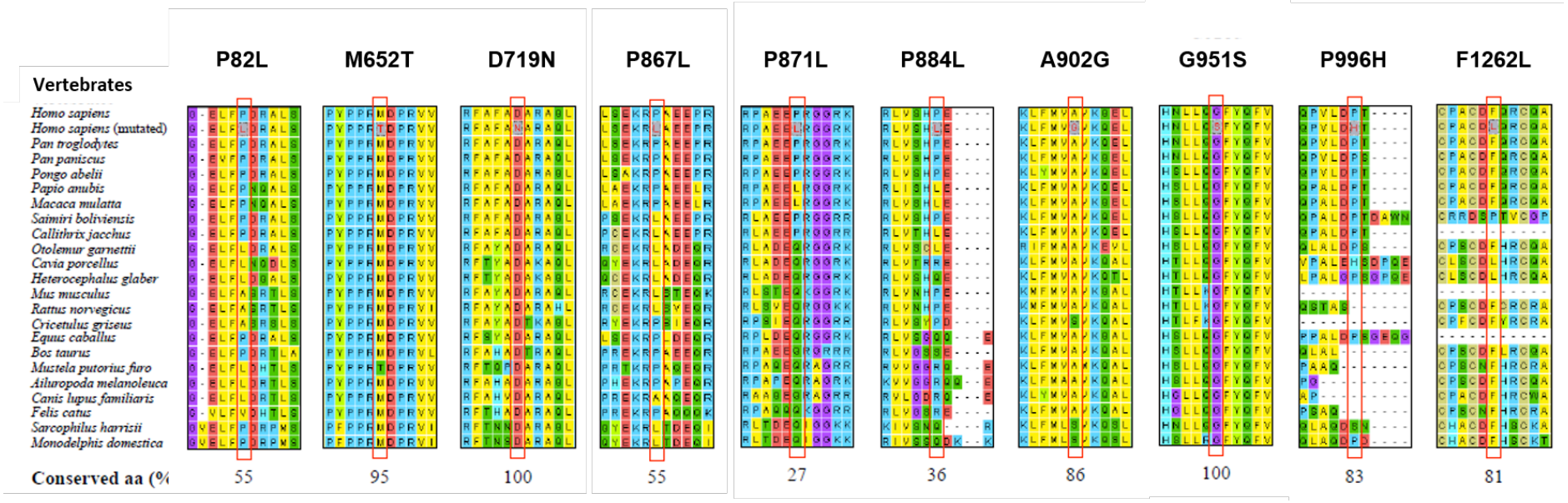


Figure 8. Amino acid alignment of RTEL1 sequences among 22 mammalian species. RTEL1-homolog amino acid sequences of 22 mammalian species available in the GenBank were aligned using the ClustalW algorithm in the MEGA tool. Substituted amino acid residues (marked with red boxes) are illustrated along with surrounding 10 residues. The second row of each panel shows a mutated version of RTEL1 amino acid sequence in Homo sapiens. Names of RTEL1 variants and mammalian species are indicated above and left the amino acid panels, respectively. RTEL1 variants are described using isoform 3 for annotation. Amino-acid gaps are due to different primary protein structures of some RTEL1 homologs. Conservation rates at individual amino acid sites (the number of species with a conserved amino acid versus the total number of species with RTEL1 sequences reported) are shown below each panel.

4.2.2. Heterozygous *RTEL1* variants in bone marrow failure

Among 8 patients with *RTEL1* variants (median age of 32 years, ranging from 6 to 63), 7 were diagnosed with AA (moderate [MAA] or severe [SAA]) and 1 with isolated neutropenia (ICUS) (Table 4). Except for NIH-1 and NIH-2, who were both biallelic (the D719N/P871L and G951S/F1262L variants, respectively), the other 6 patients carried single heterozygous *RTEL1* variants (Table 4). Three of 8 patients had very short telomeres for their age.

Patients were treated with protocols followed at NIH: androgens (danazol) for patients with telomere dysfunction or immunosuppression (IST) combined with Eltrombopag. NIH-1 responded to danazol; NIH-3 and 5 responded to IST/eltrombopag but evolved to AML/MDS or monosomy 7; NIH-4, 6, and 7 required no treatment (Table 4).

Overall, 3 of 59 patients (NIH-1, 2 and 4) had likely pathogenic heterozygous variants that were present throughout the entire *RTEL1* sequence (M652T, D719N, G951S, and F1262L); NIH-1 and 2 presented very short telomeres and telomeropathy phenotype that associated with their *RTEL1* variants. NIH-4 lacked a personal or family history consistent with a telomere syndrome. Pedigree data of NIH-1 and NIH-2 displayed variable penetrance and showed inheritance of compound heterozygosity associated with AA (Figure 9); particularly severe disease was observed in the NIH-1 proband and sibling, suggesting compound heterozygosity for the P871L and D719N variants to be particularly damaging. Isolated red blood cell macrocytosis was the sole abnormality observed in affected family member heterozygous for the G951S (Family NIH-2).

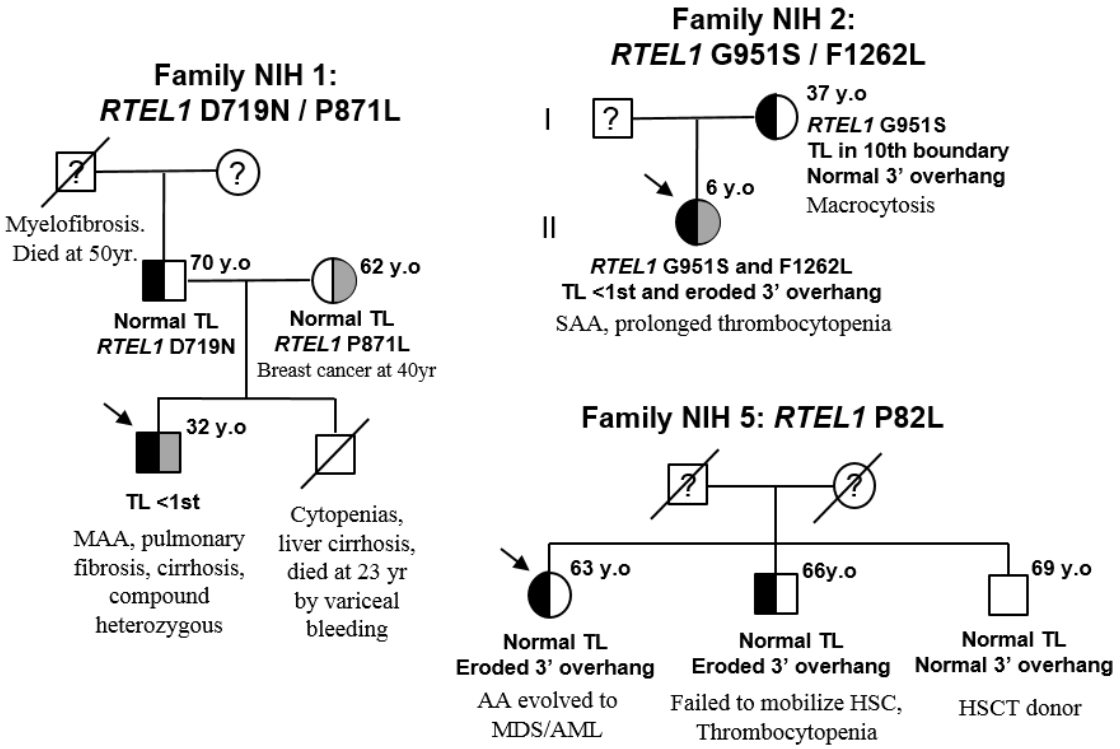


Figure 9. Pedigrees of NIH-1 and NIH-2 families identified with *RTEL1* variants. The *RTEL1* variant present in the family is described above each pedigree. Arrows indicated the proband of each family. Open circles and squares represent females and males that are non-carriers, respectively. Half or full black circles and squares represent individual who are heterozygous or homozygous for the *RTEL1* variant. Line through indicated individuals who deceased. When tested, telomere length (TL) and clinical features are described in the figure under each individual. Relatives that lack clinical and mutational data are indicated by a question mark. NIH-1 and NIH-2 were identified as a compound heterozygous (half black and gray square) as both had two *RTEL1* variants. In NIH-1, the proband's father carried the D719N (half black square) and mother carried the P871L (half grey circle) variants, both in heterozygous. In NIH-2, the father was not screened but mother carried the G951S variant in heterozygosity. Abbreviations were used as following: NIH, National Institutes of Health's cohort; yr, years; BMF, bone marrow failure; MAA, moderate aplastic anemia.

Three variants were classified as of uncertain significance and 3 as likely benign (Table 4). However, there was strong evidence of pathogenicity for 3 variants even if they failed to meet ACMG's stringent criteria. The P82L variant in NIH-5 had a supportive pedigree (Figure 9) and functional data (Figure 11) but lacked adequate in silico predictions and genetic conservation (Table 6 and Figure 8); NIH-5 evolved to MDS/AML following treatment for AA and her affected brother failed stem-cell mobilization during attempted donation for transplant. The A902G variant was well conserved (Figure 8), absent in controls (Table 6), and predicted to be damaging, but functional or pedigree data were not available (Table 6). The P871L, despite being classified as likely benign, was observed as a compound heterozygous in NIH-1 with a classical phenotype of telomeropathy.

1.1.1. RTEL1 haploinsufficiency associates with telomere shortening and single strand 3' overhang erosion independent of length

We first assessed patients' telomere integrity by measuring overall TL. As expected, patients that were biallelic presented with very short telomeres, below 1st percentile of age-matched controls. However, 6 patients had normal TL for their age (Figure 10).

We next assessed the single-stranded 3'overhangs integrity in the NIH cohort and observed that 4 of 8 patients (NIH-3, 4, 5, and 6) had eroded 3' overhangs, all of them independent of TL (Figure 11A). Despite their normal 3' overhangs, the NIH-2, NIH-7, and NIH-8 had high focal 3' overhang signal in non-denaturing blotting and no signal in the lower part of the gel, and thus, their range of 3' overhang was not as broad as healthy controls (Figure 12).

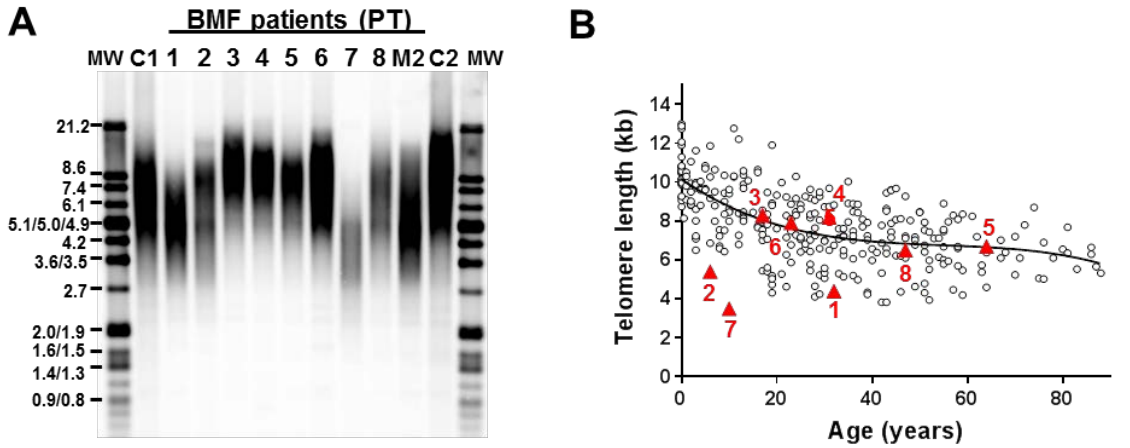


Figure 10. Telomere length measured by Southern blot (SB). (A) TL was measured in PBMCs from eight AA patients (PT), NIH-2-mother (M2), and two healthy controls (C1 and C2). Molecular weight (MW) marker sizes are indicated as kilobases (kb) on the left. **(B)** TLs of leukocytes were compared to age-matched healthy controls. Red triangles represent TLs of eight patients with *RTEL1* variants. Only NIH-1, NIH-2, and NIH-7 had very short telomeres.

Independent of length, overhang erosion associated with one likely pathogenic variant (M652T) in NIH-4 and two variants of uncertain significance and very low CADD phred score in NIH-3 and NIH-5 (P884L and P82L, respectively) (Table 4). Both NIH-3 and NIH-5 clonally evolved to monosomy 7 and MDS/AML, respectively, and presented up-regulation of genes related to DNA repair and DNA damage response. In contrast, the transcriptional profile of NIH-2 displayed upregulation of genes related to homologous recombination, apoptosis, and senescence triggered by very short telomeres (Figure 13). NIH-5 affected brother failed stem-cell mobilization during attempted donation for transplant and had short 3' overhangs whereas the non-carrier brother was healthy and the donor for bone marrow transplant (Figure 9 and 11A). Only for NIH-6, the overhang erosion associated with a likely benign *RTEL1* variant. In this case, RTEL1 may play a

secondary role in patient's disease, as the R537C *TERT* variant may play a major role in patient's telomere dysfunction.

To assess whether 3' overhang erosion was related to RTEL1 dysfunction, we measured the 3' overhangs from 14 individuals from the NIH cohort that did not carried an RTEL1 variant and in which DNA was available. The frequency of patients with eroded 3' overhang and *RTEL1* variants (50%) was higher than the frequency observed in patients without *RTEL1* variants (21%), indicating that 3' overhang shortening correlated with RTEL1 dysfunction rather than in BMF in general (Figure 11B).

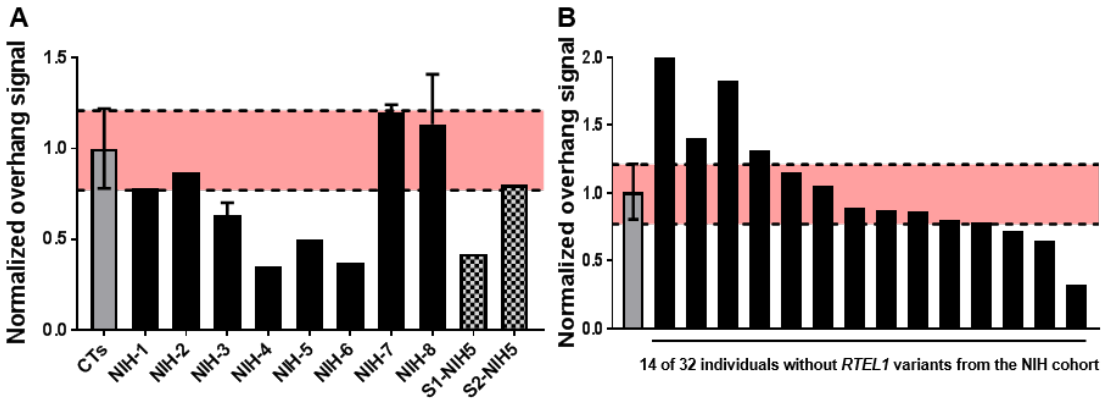


Figure 11. Normalized patients' single-stranded 3' overhang signals in comparison to controls. **(A)** Normalized single stranded 3' overhang measurement by non-denaturing SB of patients from the NIH cohort and 2 siblings of NIH-5 (S1 and S2-NIH5). Relative 3' overhangs signals were determined by normalizing the sum of chemiluminescent signal (ΣODi) from each column in the non-denaturing membrane (overhang signals) by the telomeric signal in the denatured membrane (representing total genomic DNA). Background signal (sample treated with Exo1) was subtracted from total ΣODi for all the samples. Error bars represent the standard deviation of independent experiments. Patients' 3' overhang signal were then normalized by an average of 3' overhang signals from healthy individuals used as a control in every experiment and plotted. 3' overhang signals of patients below a 95% confidence interval of 3' overhang measurements from controls (pink interval in graphic) were considered eroded. **(B)** Normalized single stranded 3' overhang measurement of 14 patients from the NIH cohort without *RTEL1* variants in which DNA was available. Three patients had telomeric overhangs below a 95% confidence interval of healthy controls and were considered short.

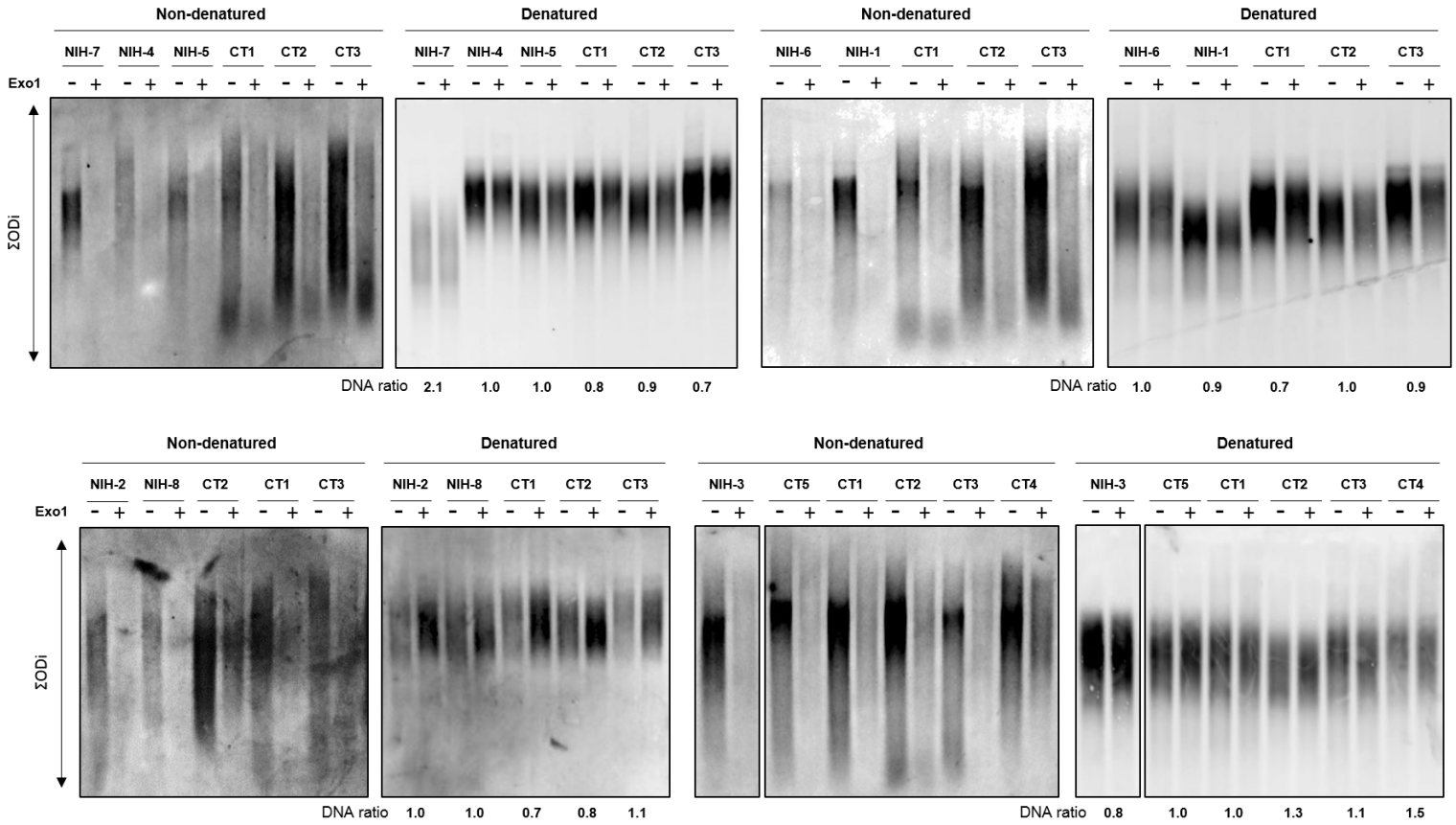


Figure 12. 3' overhang measurement by non-denaturing SB in patients from the NIH cohort. Individual lanes show overall 3'overhang signals from patients (NIH-1 to 8) and healthy controls (CT1-CT5). Smear intensities of 3' overhang signals were quantified for each sample. A relative overhang signal was determined by normalizing the sum of chemiluminescent signal (ΣODi) from each column in the non-denaturing membrane (overhang signals) by the telomeric signal in the denatured membrane (representing total genomic DNA). Normalization ratio is shown below each sample in the denatured membrane. Background signals (sample treated with Exo1) was subtracted from total ΣODi for all the samples. An average of 3' overhang signals from healthy individuals was used as a control for patient's 3' overhang normalization shown in Figure 10A-B.

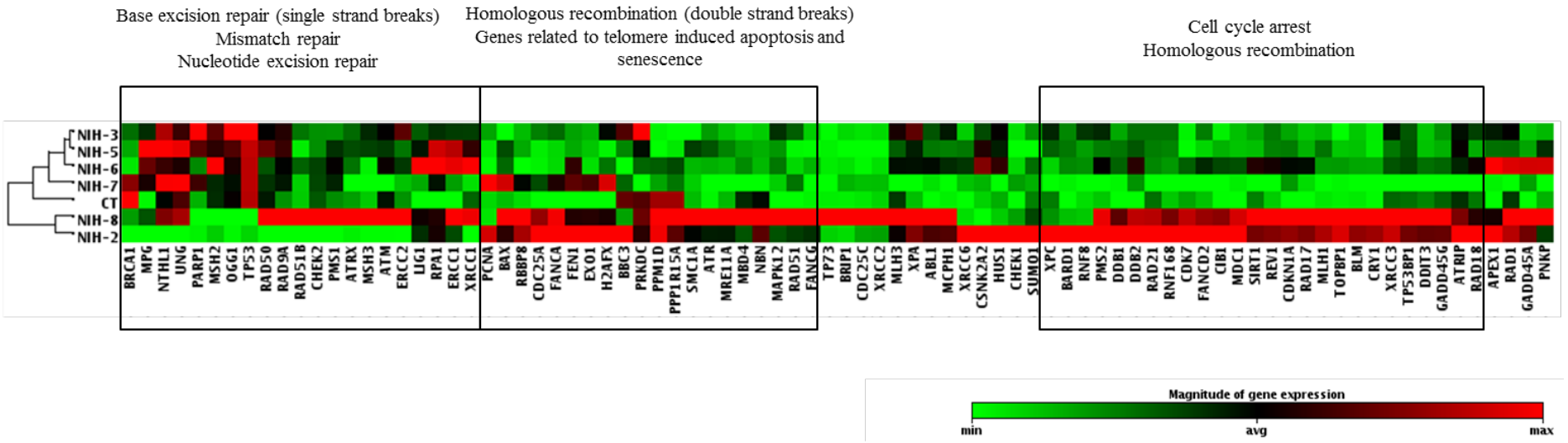


Figure 13. Clustergram of the DNA damage signaling gene expression levels in patients with *RTEL1* variants from the NIH cohort. The gene expression levels of 84 genes related with the ATM/ATR, DNA damage and Repair, Apoptosis, and Cell cycle pathways were assessed for 6 patients with *RTEL1* variants and one healthy control (CT) using the RT^2 Profiler PCR array system (Qiagen). Both NIH-3 and NIH-5 had excessive 3' overhang erosion and the genes related with DNA repair and DNA damage response (base excision repair, mismatch repair, nucleotide excision repair, and non-homologous end joining) extensively up regulated. NIH-6 had normal telomere length (TL) and 3' overhangs but gene expression pattern similar to NIH-3 and 5. Differently, in NIH-2, which had very short telomeres along with eroded 3'overhangs, the genes related with senescence and apoptosis induced by short telomeres were up regulated in comparison with the other patients. Despite NIH-8 had eroded 3' overhangs but normal TL, the gene expression pattern was similar to NIH-2. The gene expression of genes related to DNA damaging response was similar between patients with telomere shortening (NIH-2 and NIH-7) and differed from patients with eroded 3' overhangs (NIH-3 and NIH-5). The control sample was used as reference for normalization.

1.1.2. RTEL1 modulates TRF2 expression in 293T cells

To assess the impact of *RTEL1* variants in the helicase interaction with TRF2, we generated 293T cells stably expressing the RTEL1-FLAG WT, the F1262L, P82L, M652T, D719N, and G951S variants.

In bulk 293T-RTEL1-FLAG cell lines, we first evaluated recombinant RTEL1-FLAG and TRF2 expression (Figure 14A). Surprisingly, TRF2 expression was increased in the WT and the M652T-293T cell lines when compared to the 293T-empty vector, but it was significantly downregulated in 293T-F1262L cells (Figure 14A left panel). To validate our results, we examined both FLAG and TRF2 expression in a single-cell isolated clone of 293T-RTEL1-FLAG and immunoprecipitated RTEL1-FLAG to check its binding with TRF2.

TRF2 expression was increased in cells stably expressing exogenous WT RTEL1 isoform 3 and the variants M652T, D719N and G951S. The 293T-P82L cell had a 4-fold increase in TRF2 expression compared to WT (Figure 14B; Input panel and Figure 14A). As seen with bulk cells, the TRF2 expression in the single-cell isolated clone of 293T-F1262L cells were significantly downregulated as compared to WT or the control (Figure 14A; Input panel). In our patients, neither NIH-2 displayed TRF2 downregulation nor NIH-5 displayed TRF2 upregulation in peripheral blood cells assessed by PCR array (Figure 14C). The RTEL1 and TRF2 interaction was preserved in WT and 293T cells carrying *RTEL1* variants (Figure 14A; pulldown). The ratio of TRF2 immunoprecipitated with FLAG was similar between WT and mutants. The lowest ratio of TRF2 immunoprecipitated was in 293T-F1262L cells. In summary, the F1262L RING domain variant did not completely disrupted the RTEL1 and TRF2 interaction, but TRF2 downregulation perturbed the RTEL1 interaction with this shelterin.

1.1.3. RTEL1 patients display a lower t-circles amount than controls

RTEL1 deficiency are linked to aberrant accumulation of t-circles in cells (Sarek *et al.*, 2015). We checked t-circles formation in the PB samples from NIH patients by the TC assay. We did not observe any t-circles accumulation in patients' cells, instead, all of them presented lower t-circle amounts when compared to the asymptomatic mother of NIH-2 (P1-NIH-2) and a healthy control (CT1). Among patients, biallelic patients were found with more t-circles than heterozygous *RTEL1* patients (Figure 14D).

1.1.4. Other germline variants

Three patients also carried variants in other telomere-associated genes: *TERT, TERC,* and *SLX4* (Table 4). A novel *TERT* R537C variant was identified in NIH-6 concomitant with the *RTEL1* P867L. The *TERC* r.287 C>A identified in NIH-7, who had moderate AA, early hair graying, and frequent miscarriages, was previously reported and predicted to be pathogenic (Vulliamy *et al.*, 2011). NIH-7 had very short telomeres but not 3' overhang erosion. Additionally, we identified a variant in the *SLX4* gene in NIH-4 that encodes an endonuclease that participates in t-loop excision and telomere shortening in RTEL1 deficient cells (Vannier *et al.*, 2012). SLX4 loss of function rescues the telomere loss phenotype *in vitro* (Vannier *et al.*, 2012), and its haploinsufficiency can account for normal TL even when RTEL1 is impaired.

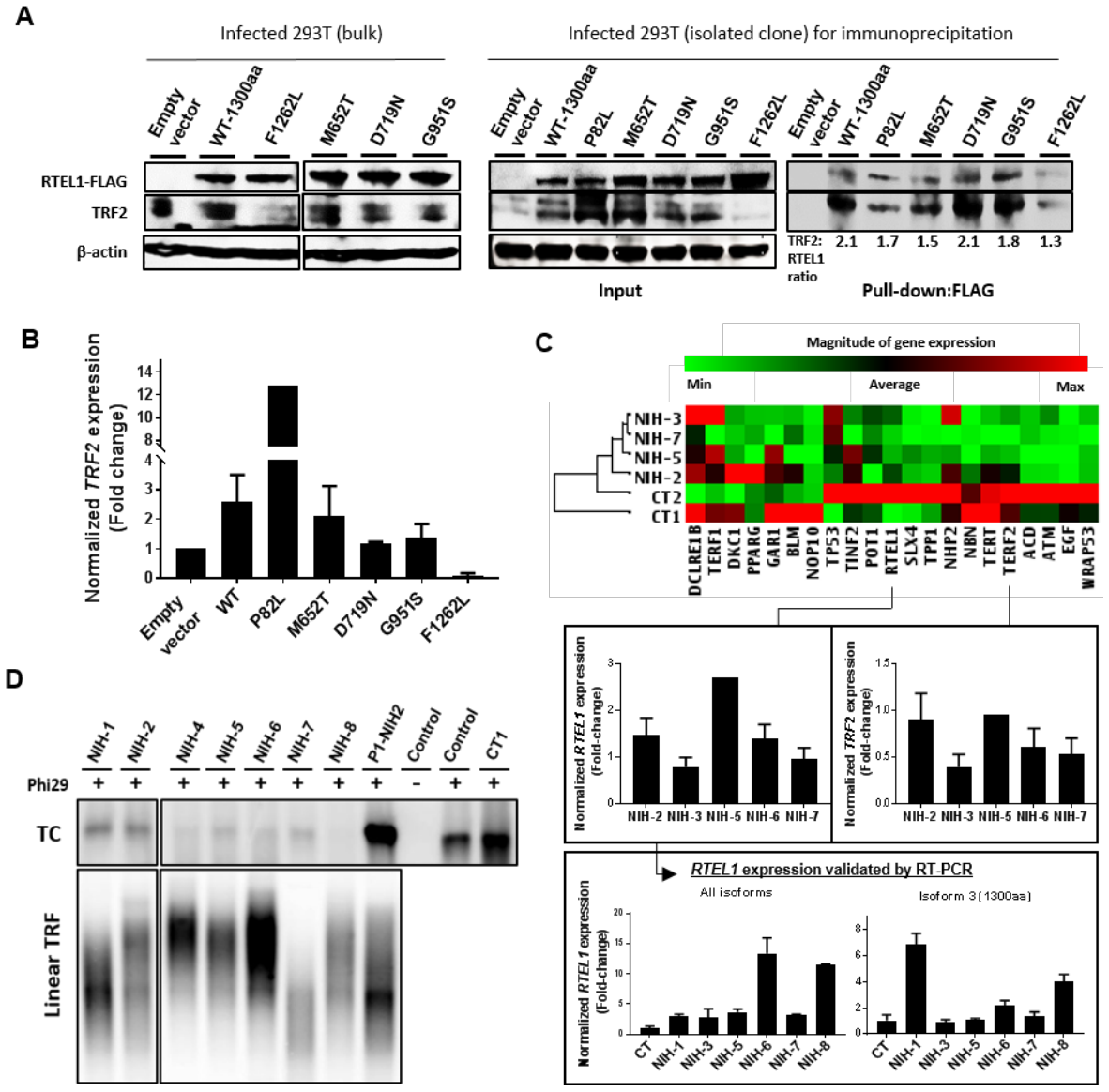


Figure 14. Impact of *RTEL1* variants in telomere maintenance. (A) Western blot analysis. Left panel: Whole extracts of bulk 293T cells stably expressing recombinant RTEL1-FLAG WT or one of the following RTEL1 variants: M652T, D719N, G951S, and F1262L. 293T infected with an empty vector was used as a control. Right panel: Whole extracts (input) of isolated RTEL1-FLAG WT or 293T clones with the P82L, M652T, D719N, G951S, and F1262L variants were immunoprecipitated with anti-FLAG to evaluate RTEL1 and TRF2 interaction. Proteins expression was analyzed with antibodies as indicated. **(B)** TRF2 expression in both bulk infected 293T cells and isolated clones normalized by the control (empty vector). Error bars represent the standard deviation between TRF2 expression in bulk 293T and isolated clones. **(C)** Clustergram of the telomere biology gene expression levels in 4 patients with *RTEL1* variants from the NIH cohort and 2 controls using the RT2 Profiler PCR array system (Qiagen). Control samples were used as reference for normalization. For comparison, the RTEL1 and TRF2 fold change expression relative to controls were plotted and shown in the graphic below the heatmap. The *RTEL1* gene expression was validated by RT-PCR using a Taqman® probe that detected the boundary of exon 7-8 in all isoforms and another probe that detected the RING domain in isoform 3. **(D)** Phi29 dependent T-circle amplification assay in patients' peripheral blood. T-circles (TC) were detected in lower amount in patients with heterozygous variants compared to biallelic. The asymptomatic mother of NIH-2 (P1-NIH2) and a healthy control (CT1) presented the highest amount of t-circles. DNA extracted from VA13 cells was used as control of the assay.

DISCUSSION

2. DISCUSSION

We report novel germline variants associated with acquired or constitutional AA in two independent cohorts. The phenotype of patients with acquired and constitutional AA overlaps and differential diagnosis of these diseases rely on a combination of laboratory tests, clinical features, and family history. TL measurement is commonly used as the first line tool in constitutional BMF investigation, as patients with telomeropathies classically present telomere shortening for their age (Calado and Young, 2008; Gutierrez-Rodrigues *et al.*, 2014).

However, in both FMRP and NIH cohorts, we identified patients with AA and normal TL carrying germline variants in helicase genes; a variant in the *DHX36* gene (patient NGS_4 from FMRP cohort) and the heterozygous *RTEL1* variants in NGS_2 (FMRP cohort), and NIH-4 to 6 (NIH cohort). Here, independently of the AA etiology, acquired or constitutional, the genetic screening provided additional data critical for patients' treatment.

Both exome and targeting sequencing have been used to characterize the spectrum of BMF diseases. The distinction between constitutional and acquired BMF is a concern of physicians, as it has a direct impact on patients' management. The importance of a diagnosis for constitutional/inherited BMF disorder is often underestimated. First, because IST is usually ineffective in patients with constitutional BMF, even though patients with mutations in telomere biology genes respond to IST (Townsley *et al.*, 2014). In our study, patients without functional relevant germline variants responded to IST, while patients with under-reported variants in genes related to telomere maintenance had poor outcomes. Also, there are implications for HSCT donor selection and choice of transplant conditioning regimen. Family screening is imperative to avoid using affected 'silent' siblings. Finally, the risk of later cancers is higher in constitutional compared to acquire BMF (Bennett and Orazi, 2009). Indeed, four patients carrying variants in *DHX36* (NGS_4), *TERT* (NGS_7), and *RTEL1* (NIH-3 and 5)

clonally evolved to AML/MDS or monosomy 7 after IST, even when a hematologic response was achieved.

One of the major challenges in genetic studies is to ascertain the pathogenicity of a given variant in order to assess its role in disease development and progression. The contribution of a variable clonal hematopoiesis rather than germline variants hampers the understanding of the role of genetic variants in AA. Most of the studies associate the response and outcomes of patients with acquired AA with the presence of somatic clones in peripheral blood (Dumitriu *et al.*, 2015; Yoshizato *et al.*, 2015). Albeit, it was found that somatic clones with an allele frequency of 0.02% in genes related to myeloid malignancies are present in 95% of healthy individuals over 60 years old (Young *et al.*, 2016). If somatic clones are commonly acquired with aging, what triggers the clonal expansion of this cells and how these clones are associated with the development of hematologic diseases such as AA? Also, it is not known whether germline variants in genes related to hematologic diseases could also be related to responsiveness to IST or to the heterogeneity of clinical features seen in AA patients. These are the questions that have been studied by different researchers and stands as an extensive field for further investigation.

In our cohort of acquired AA, we found no variants, somatic or germline, in genes related to clonal hematopoiesis such as *DMNT3A*, *ASXL1*, and *BCOR* (Yoshizato *et al.*, 2015). However, in our study, the mean coverage of our samples might be too low to detect clones in small frequencies with confidence and the error correcting sequencing approach (Young *et al.*, 2016) might be required to improve our data.

Since AA is a very heterogeneous disease and is not associated with a classical Mendelian inheritance, the mutational profile of germline variants in patients with no clinical and laboratorial features of a constitutional disease was not comprehensively addressed in past studies. Our study highlights the importance of genetic screening for patients with BMF with

normal TL, without a family history or clinical stigmata. Corroborating our study, germline mutations in BMF related genes were identified in a subset of pediatric and young adults referred for HSCT for AA and MDS between 1990-2012, in which sample was available in a biorepository (Keel *et al.*, 2016). They identified a potentially damaging variant in 5.1% (5/98) of AA patients and 13.6% (15/110) of MDS patients treated as having an acquired disease.

In BMF, a targeting sequencing is preferred to screen patients, since the genes frequently mutated in BMF are known, and exome sequencing is costly as well as requires a robust strategy to filter relevant variants. The data analysis from the cohort with acquired AA did not reveal a clear genetic pattern associated with their IST response; most of the variants were of uncertain significance and the small sample size hamper further conclusions. Only the patient NGS_04 was found with a variant in the helicase gene *DHX36* that might be involved in his disease. In contrast, the sequencing of the cohort selected based on their suspicion to have an inherited disease identified a molecular defect that might be pathogenic in up to 40% of patients, including the *RTEL1* variants. This rate could have been higher if we had screened our constitutional AA patients with the expanded targeting sequencing panel instead of the commercial one (only with 49 genes from Chicago Laboratories). The customized targeting sequencing panel comprised 165 genes linked to hematologic diseases or that are potentially targets due to their involvement to signaling pathways commonly affected in BMF. Then, this panel expanded the probability to identify pathogenic variants related to telomere diseases or hematologic malignancies. Several genes have been added to the commercial panel from Chicago Laboratories, but this panel remains strictly clinical and covers only the genes identified in BMF patients.

In our study, we identified novel variants in the helicases genes *RTEL1* and *DHX36* involved in telomere maintenance. Although *RTEL1* have been linked to telomeropathies (Le Guen *et al.*, 2013; Walne *et al.*, 2013), the *DHX36* appears as a new mutated gene in telomere

diseases. The protein DHX36 is part of the DEAH-box helicases genes that are not commonly screened in patients with suspected telomeropathies. The mechanism by which DHX36 interacts with telomeres has never been reported and there are no reports of patients with mutations in this gene. Recently, the DEAH-box helicase *DDX41* was included in the list of genes related to inherited BMF, as it was identified mutated in a subset of familial MDS and AML.

Biallelic mutations in *RTEL1*, either homozygous or compound heterozygous have been extensively linked to DC and Hoyeraal-Hreidarsson (HH) syndrome, severe forms of DC presenting in childhood and with extremely short telomeres (Ballew *et al.*, 2013; Deng *et al.*, 2013; Le Guen *et al.*, 2013; Walne *et al.*, 2013; Touzot and Kermasson, 2016). In contrast to heterozygous mutations in *TERC, TERT,* and *TINF2,* that have been identified in up to 10% of patients with AA, MDS, and AML (Yamaguchi *et al.*, 2003; Yamaguchi *et al.*, 2005; Calado, Regal, Hills, *et al.*, 2009), the role of heterozygous *RTEL1* variants in BMF have not been addressed.

For our knowledge, this is the first report that associated heterozygous *RTEL1* variants classified as likely pathogenic with AA in both at an early age and in adulthood. We classified variants using a combination of evidence that stringently assessed the pathogenicity of *RTEL1* variants. We identified 3 likely pathogenic *RTEL1* variants (all of them were novel) in 3 unrelated patients. Likely pathogenic *RTEL1* variants were enriched in patients selected as possibly having inherited BMF as compared to ExAC database. In contrast to biallelic telomeropathy variants, RTEL1 haploinsufficiency associated with clinical manifestations that did not meet criteria for any specific constitutional BMF syndromes, such as DC and HH syndrome. Constitutional BMF, especially in adults, is likely under-reported, due to incomplete penetrance, reduced expressivity and disease anticipation, as well as a lack of awareness by clinicians that it may present without mucocutaneous features but with pulmonary fibrosis or

cirrhosis (Collopy *et al.*, 2015). The pathogenicity of heterozygous *RTEL1* variants must be evaluated with caution since many very rare variants are present in population databases. In 3 patients, the *RTEL1* variant (M344T, P891L, and P996H) was not etiologic. The number of patients in which *RTEL1* variants are not pathogenic may be higher, as 3 variants were of uncertain significance.

As seen with other telomere biology components, heterozygous *RTEL1* variants had a variable penetrance, possibly acting as disease modulators in a complex genomic architecture. Penetrance of *TERT* or *TERC* mutations is very heterogeneous (Winkler *et al.*, 2013) and other genetic or epigenetic factors have been hypothesized to modulate disease manifestations. "Epigenetic-like" inheritance of short TL can result in human disease in the absence of a telomerase mutation (Xin *et al.*, 2007). Biallelic *TERT* variants were shown to aggravate patient phenotype (Marrone *et al.*, 2007), and tri-allelic inheritance of homozygous *TERT* and heterozygous *TERC* variant was seen in a family member with severe DC phenotype (Collopy *et al.*, 2015). RTEL1 haploinsufficiency was not affected by downregulation of the helicase expression in patients' cells as *RTEL1* expression seen in patients carrying heterozygous variants was higher or similar to controls (Figure 14C). Then, it is unlikely that an epigenetic silencing, aberrant splicing or deletion of the normal *RTEL1* allele would be causing a recessive phenotype along with the heterozygous variant.

We identified a co-occurrence of *RTEL1* and *TERT/TERC/SLX4* variants in three patients. However, *RTEL1* co-existing variants were not associated with a more aggressive hematologic disease. One patient with a likely benign *RTEL1* variant and the R537C *TERT* variant had a coexisting PNH clone (NIH-6; Table 5), which opposes to the pathogenic role of her variants. In this patient, the *RTEL1* variant was classified as likely benign, but the *TERT* variant remains as potentially pathogenic. BMF patients with mutations in telomere biology genes may present a PNH clone suggesting a coexisting immune component contributing to

marrow failure.

Surprisingly, the *RTEL1* variants associated with eroded 3' overhang independent of TL. Eroded 3' overhangs are markers of cell senescence and telomere dysfunction (Stewart *et al.*, 2003), and have been described in AA patients with *TERT/TERC* mutations (Calado, Regal, Kajigaya, *et al.*, 2009). RTEL1 haploinsufficiency was only associated with overhang shortening in one previous paper; two siblings with HH syndrome and compound heterozygous with R974X and M492I variants presented 3' overhang erosion, but differently from our study, they had very short telomeres. Their heterozygous parents, although asymptomatic, also presented with short telomeres and 3' overhang erosion (Deng *et al.*, 2013). In primary fibroblasts of other HH family, excessive 3' overhang erosion appeared to impair cell proliferation and to promote extensive DNA damage response (DDR) activation independent of TL (Lamm *et al.*, 2009). The only component found to be mutated in a patient with TL-independent telomere dysfunction was the Apollo enzyme (Touzot *et al.*, 2010). As RTEL1, the Apollo/SNM1B nuclease is required for telomere-end processing and replication. Apollo regulates 3' overhang length and directly binds to TRF2 to maintain telomere homeostasis during cell replication in the S phase (Wu *et al.*, 2010). Disturbance in the telomere replication and protection pathways, of which Apollo and RTEL1 are a part, may lead to telomere dysfunction independent of telomere shortening. Two patients with eroded 3' overhangs and normal TL evolved to AML/MDS after IST treatment. Inability to suppress DNA-damage response combined with normal TL may predispose cells to genomic instability and leukemogenesis, rather than to drive pathways of senescence or apoptosis, as occurs in cells with critically short telomeres (Figure 15). Indeed, expression levels of DNA damage-related genes were different between patients with eroded 3' overhangs and short telomeres (Figure 13). Also, MDS and AML were the primary diagnoses of 2 patients identified with variants that had strong evidence for pathogenicity, despite classified as of uncertain significance.

However, we infer from our data that *RTEL1* variants classified as likely pathogenic were not solely implicated in excessive 3' overhang erosion and inappropriate telomere capping.

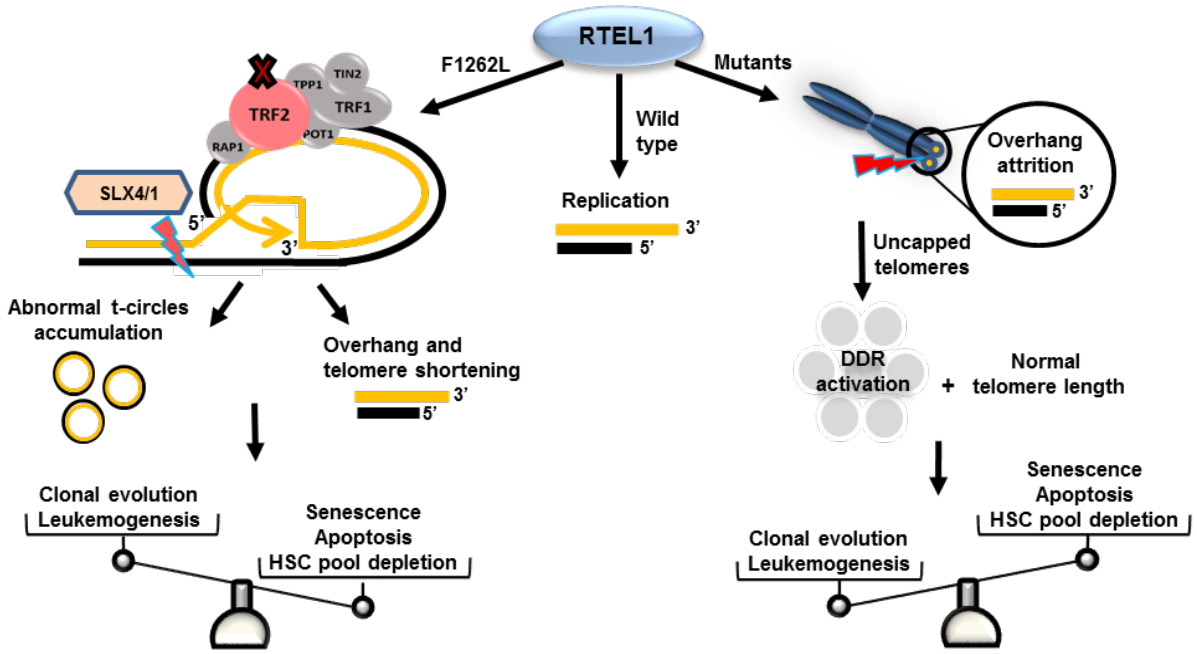


Figure 15. Schematic representation of dual *RTEL1* roles in telomere dysfunction. In normal conditions, RTEL1 promotes G4 quadruplex and t-loop unwinding for DNA replication. The F1262L mutation inhibits RTEL1-TRF2 interaction and then t-loop resolution. Thus, abnormal SXL4/1 activation cuts persistently t-loops, leading to 3' overhang and telomere shortening as well as the accumulation of t-circles in cells. Short telomeres, commonly seen in patients with mutated RTEL1, trigger cell senescence and apoptosis via p53, leading to hematopoietic stem cell depletion in BM. However, some RTEL1 mutations are related to excessive 3' overhang attrition in absence of telomere shortening. In a context that cells maintain their TLs, the sustained DDR activation caused by uncapped telomeres may promote leukemogenesis by molecular pathways distinct from typical accelerated telomere attrition associated with very short telomeres.

Assessment of the impact of variants in gene function and the use of frequency controls are critical. Pathogenicity assessment for the heterozygous *RTEL1* gene is challenging, as approximately 1 of 100 individuals in the general population carries a rare variant in this gene with CADD score over 15 (Figure 7B). Consequently, detailed extended family history and functional data were imperative to classify heterozygous *RTEL1* variants as pathogenic. It is important to point out that there is no single assay to assess the functional meaning of variants

located outside the RING domain. The t-circles assay is used as a marker to identify RTEL1 deficient cells; however, as seen in our results, other studies did not identify any accumulation of t-circles in patients' cells (Deng *et al.*, 2013; Touzot and Kermasson, 2016). The interpretation of functional assays is still limited to what is known about the role of RTEL1 dysfunction in cell biology. Here, we described RTEL1 as a modulator of TRF2 expression in vitro. Thus, TRF2 expression regulation may be an alternative mechanism related to RTEL1 dysfunction rather than impaired T-loop disassembly alone.

Our study is limited by sample size and shows primitive results that hamper the association between patients' genetic profile and response to treatment. The genetic screening of patients suspected to have acquired AA due to their normal TLs revealed a genetic factor associated with IST response. Also, in a cohort consecutively screened for germline mutations, heterozygous *RTEL1* variants associated with both overall telomere shortening and single-stranded 3' overhang erosion independent of length. Then, TL assessment alone may not identify all primary telomere defects.

In the high-throughput NGS era, we have more questions than answers. It is entirely required to have a consolidated strategy to analyze the sequencing data to allow reproducibility and confidence in the results. However, there are a variety of pipelines for data analysis and researchers need to choose which parameters are the priority for their study and what is the minimum quality data acceptable to call variants. In our pipeline, we applied hard filters to hid off low-quality variants but we accepted the risk of discarding potentially damaging variants. Also, we filtered out variants with a frequency higher than 0.1% in genome databases, with no impact on gene functions (non-synonymous or intronic variants) and with a low number of reads. The variants called were then evaluated for their pathogenicity and their involvement in patients' disease. We first assessed the role of variants identified in genes already linked to BMF syndromes. Then, the variants predicted in silico as highly damaging and in genes critical

to biological processes. In silico predictions are limited in their ability to predict pathogenicity of rare variants. From a practical standpoint, the NGS approach identifies many rare variants whose pathogenicity will remain unclear in the absence of further analysis**.** Compared to Sanger sequencing, the NGS can screen a much higher number of patients and genes in a lower time frame with a low cost. However, unfortunately, some factors still limit the use of NGS in Brazil: (1) we lack human resources specialized in data analysis. A computational team is necessary to work together with clinicians and staff scientists in data analysis process; (2) it is not economically feasible to sequencing in our national facilities. Due to our bureaucratic system, reagents are more expensive, the cost is higher than sequencing in facilities outside the country, and the time to get results is unreasonable longer than the expected; (3) We lack a federal politics to support and to funding the NGS sequencing projects, that despite being cheaper than Sanger need a high investment at once. The chaotic political and economic structure in which Brazilian science is part are even worse with the current federal administration, that continually chooses to underestimate the value of the science in public health and the country development. Despite discouraging and unlikely to be able to follow the international scientific pace, it is possible to offer a service of quality and highly specialized for our patients and contribute to the national science. Here, we provided data from a pioneer sequencing project that screened patients with BMF.

The combination of different tools for *in silico* prediction, well-established criteria to classify variants, a variant allele frequency control from ExAC, familial investigation, and functional data were critical to increase the confidence of our analysis and to provide evidence that variants classified as likely pathogenic were etiologic. We classified 3 variants as pathogenic and extended the spectrum of disease related to RTEL1 dysfunction, which contributes to a better understanding of the biology and mechanisms related to telomere diseases as well as to a more appropriated management of those patients.

REFERENCES

6. REFERENCES

ANDERSON, B. H. et al. Mutations in CTC1, encoding conserved telomere maintenance component 1, cause Coats plus. **Nat Genet,** v. 44, n. 3, p. 338-42, Mar 2012. ISSN 1546-1718. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/22267198 >.

BAERLOCHER, G. M. et al. Flow cytometry and FISH to measure the average length of telomeres (flow FISH). **Nat Protoc,** v. 1, n. 5, p. 2365-76, 2006. ISSN 1750-2799. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/17406480 >.

BALLEW, B. J. et al. Germline mutations of regulator of telomere elongation helicase 1, RTEL1, in Dyskeratosis congenita. **Hum Genet,** v. 132, n. 4, p. 473-80, Apr 2013. ISSN 1432-1203. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/23329068 >.

BENNETT, J. M.; ORAZI, A. Diagnostic criteria to distinguish hypocellular acute myeloid leukemia from hypocellular myelodysplastic syndromes and aplastic anemia: recommendations for a standardized approach. **Haematologica,** v. 94, n. 2, p. 264-8, Feb 2009. ISSN 1592-8721. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/19144661 >.

BICK, D.; DIMMOCK, D. Whole exome and whole genome sequencing. **Curr Opin Pediatr,** v. 23, n. 6, p. 594-600, Dec 2011. ISSN 1531-698X. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/21881504 >.

BOOY, E. P. et al. The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. **Nucleic Acids Res,** v. 40, n. 9, p. 4110-24, May 2012. ISSN 1362-4962. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/22238380 >.

CALADO, R. T. et al. Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. **Proc Natl Acad Sci U S A,** v. 106, n. 4, p. 1187-92, Jan 2009. ISSN 1091-6490. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/19147845 >.

CALADO, R. T. et al. Erosion of telomeric single-stranded overhang in patients with aplastic anaemia carrying telomerase complex mutations. **Eur J Clin Invest,** v. 39, n. 11, p. 1025-32, Nov 2009. ISSN 1365-2362. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/19674077 >.

CALADO, R. T.; YOUNG, N. S. Telomere maintenance and human bone marrow failure. **Blood,** v. 111, n. 9, p. 4446-55, May 2008. ISSN 1528-0020. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/18239083 >.

CALADO, R. T.; YOUNG, N. S. Telomere diseases. **N Engl J Med,** v. 361, n. 24, p. 2353-65, Dec 2009. ISSN 1533-4406. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/20007561 >.

CLÉ, D. V. et al. Repeat course of rabbit antithymocyte globulin as salvage following initial therapy with rabbit antithymocyte globulin in acquired aplastic anemia. **Haematologica,** v. 100, n. 9, p. e345-7, Sep 2015. ISSN 1592-8721. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/25862703 >.

COLLINS, K.; MITCHELL, J. R. Telomerase in the human organism. **Oncogene,** v. 21, n. 4, p. 564-79, Jan 2002. ISSN 0950-9232. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/11850781 >.

COLLOPY, L. C. et al. Triallelic and epigenetic-like inheritance in human disorders of telomerase. **Blood,** v. 126, n. 2, p. 176-84, Jul 2015. ISSN 1528-0020. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/26024875 >.

DENG, Z. et al. Inherited mutations in the helicase RTEL1 cause telomere dysfunction and Hoyeraal-Hreidarsson syndrome. **Proc Natl Acad Sci U S A,** v. 110, n. 36, p. E3408-16, Sep 2013. ISSN 1091-6490. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/23959892 >.

DESMOND, R. et al. Eltrombopag restores trilineage hematopoiesis in refractory severe aplastic anemia that can be sustained on discontinuation of drug. **Blood,** v. 123, n. 12, p. 1818-25, Mar 2014. ISSN 1528-0020. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/24345753 >.

DUMITRIU, B. et al. Telomere attrition and candidate gene mutations preceding monosomy 7 in aplastic anemia. **Blood,** v. 125, n. 4, p. 706-9, Jan 2015. ISSN 1528-0020. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/25406353 >.

GUTIERREZ-RODRIGUES, F. et al. Direct comparison of flow-FISH and qPCR as diagnostic tests for telomere length measurement in humans. **PLoS One,** v. 9, n. 11, p. e113747, 2014. ISSN 1932-6203. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/25409313 >.

HENSON, J. D. et al. DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. **Nat Biotechnol,** v. 27, n. 12, p. 1181-5, Dec 2009. ISSN 1546-1696. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/19935656 >.

HONG, S. H.; YANG, Z.; PRIVALSKY, M. L. Arsenic trioxide is a potent inhibitor of the interaction of SMRT corepressor with Its transcription factor partners, including the PML-retinoic acid receptor alpha oncoprotein found in human acute promyelocytic leukemia. **Mol Cell Biol,** v. 21, n. 21, p. 7172-82, Nov 2001. ISSN 0270-7306. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/11585900 >.

HSU, A. P. et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. **Blood,** v. 121, n. 19, p. 3830-7, S1-7, May 2013. ISSN 1528-0020. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/23502222 >.

JONES, W. D. et al. De novo mutations in MLL cause Wiedemann-Steiner syndrome. **Am J Hum Genet,** v. 91, n. 2, p. 358-64, Aug 2012. ISSN 1537-6605. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/22795537 >.

KANNENGIESSER, C. et al. Heterozygous RTEL1 mutations are associated with familial pulmonary fibrosis. **Eur Respir J,** v. 46, n. 2, p. 474-85, Aug 2015. ISSN 1399-3003. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/26022962 >.

KAPOOR, V.; TELFORD, W. G. Telomere length measurement by fluorescence in situ hybridization and flow cytometry. **Methods Mol Biol,** v. 263, p. 385-98, 2004. ISSN 1064-3745. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/14976379 >.

KEEL, S. B. et al. Genetic features of myelodysplastic syndrome and aplastic anemia in pediatric and young adult patients. **Haematologica,** v. 101, n. 11, p. 1343-1350, Nov 2016. ISSN 1592-8721. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/27418648 >.

KHINCHA, P. P.; SAVAGE, S. A. Genomic characterization of the inherited bone marrow failure syndromes. **Semin Hematol,** v. 50, n. 4, p. 333-47, Oct 2013. ISSN 1532-8686. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/24246701 >.

KILLICK, S. B. et al. Guidelines for the diagnosis and management of adult aplastic anaemia. **Br J Haematol,** v. 172, n. 2, p. 187-207, Jan 2016. ISSN 1365-2141. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/26568159 >.

KIRCHER, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. **Nat Genet,** v. 46, n. 3, p. 310-5, Mar 2014. ISSN 1546-1718. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/24487276 >.

KNIGHT, S. W. et al. X-linked dyskeratosis congenita is predominantly caused by missense mutations in the DKC1 gene. **Am J Hum Genet,** v. 65, n. 1, p. 50-8, Jul 1999. ISSN 0002-9297. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/10364516 >.

LAMM, N. et al. Diminished telomeric 3' overhangs are associated with telomere dysfunction in Hoyeraal-Hreidarsson syndrome. **PLoS One,** v. 4, n. 5, p. e5666, 2009. ISSN 1932-6203. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/19461895 >.

LATTMANN, S. et al. The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. **Nucleic Acids Res,** v. 39, n. 21, p. 9390-404, Nov 2011. ISSN 1362-4962. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/21846770 >.

LE GUEN, T. et al. Human RTEL1 deficiency causes Hoyeraal-Hreidarsson syndrome with short telomeres and genome instability. **Hum Mol Genet,** v. 22, n. 16, p. 3239-49, Aug 2013. ISSN 1460-2083. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/23591994 >.

MARRONE, A. et al. Telomerase reverse-transcriptase homozygous mutations in autosomal recessive dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome. **Blood,** v. 110, n. 13, p. 4198-205, Dec 2007. ISSN 0006-4971. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/17785587 >.

MARSH, J. C. et al. Guidelines for the diagnosis and management of aplastic anaemia. **Br J Haematol,** v. 147, n. 1, p. 43-70, Oct 2009. ISSN 1365-2141. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/19673883 >.

MARSH, J. C.; KULASEKARARAJ, A. G. Management of the refractory aplastic anemia patient: what are the options? **Hematology Am Soc Hematol Educ Program,** v. 2013, p. 87-94, 2013. ISSN 1520-4383. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/24319168 >.

MOYZIS, R. K. et al. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. **Proc Natl Acad Sci U S A,** v. 85, n. 18, p. 6622-6, Sep 1988. ISSN 0027-8424. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/3413114 >.

PAIVA, R. M.; CALADO, R. T. Telomere dysfunction and hematologic disorders. **Prog Mol Biol Transl Sci,** v. 125, p. 133-57, 2014. ISSN 1878-0814. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/24993701 >.

PALM, W.; DE LANGE, T. How shelterin protects mammalian telomeres. **Annu Rev Genet,** v. 42, p. 301-34, 2008. ISSN 0066-4197. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/18680434 >.

PAREEK, C. S.; SMOCZYNSKI, R.; TRETYN, A. Sequencing technologies and genome sequencing. **J Appl Genet,** v. 52, n. 4, p. 413-35, Nov 2011. ISSN 2190-3883. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/21698376 >.

RAO, R. C.; DOU, Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. **Nat Rev Cancer,** v. 15, n. 6, p. 334-46, Jun 2015. ISSN 1474-1768. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/25998713 >.

RICHARDS, S. et al. 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. **Genet Med,** v. 17, n. 5, p. 405-24, May 2015. ISSN 1530-0366. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/25741868 >.

SAREK, G. et al. TRF2 recruits RTEL1 to telomeres in S phase to promote t-loop unwinding. **Mol Cell,** v. 57, n. 4, p. 622-35, Feb 2015. ISSN 1097-4164. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/25620558 >.

SAVAGE, S. A. et al. TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita. **Am J Hum Genet,** v. 82, n. 2, p. 501-9, Feb 2008. ISSN 1537-6605. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/18252230 >.

SCHEINBERG, P. et al. Horse versus rabbit antithymocyte globulin in acquired aplastic anemia. **N Engl J Med,** v. 365, n. 5, p. 430-8, Aug 2011. ISSN 1533-4406. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/21812672 >.

SCHEINBERG, P.; YOUNG, N. S. How I treat acquired aplastic anemia. **Blood,** v. 120, n. 6, p. 1185-96, Aug 2012. ISSN 1528-0020. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/22517900 >.

SCHERTZER, M. et al. Human regulator of telomere elongation helicase 1 (RTEL1) is required for the nuclear and cytoplasmic trafficking of pre-U2 RNA. **Nucleic Acids Res,** v. 43, n. 3, p. 1834-47, Feb 2015. ISSN 1362-4962. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/25628358 >.

SEXTON, A. N.; COLLINS, K. The 5' guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. **Mol Cell Biol,** v. 31, n. 4, p. 736-43, Feb 2011. ISSN 1098-5549. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/21149580 >.

SMALDINO, P. J. et al. Mutational Dissection of Telomeric DNA Binding Requirements of G4 Resolvase 1 Shows that G4-Structure and Certain 3'-Tail Sequences Are Sufficient for Tight and Complete Binding. **PLoS One,** v. 10, n. 7, p. e0132668, 2015. ISSN 1932-6203. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/26172836 >.

STEWART, S. A. et al. Erosion of the telomeric single-strand overhang at replicative senescence. **Nat Genet,** v. 33, n. 4, p. 492-6, Apr 2003. ISSN 1061-4036. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/12652299 >.

STUART, B. D. et al. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. **Nat Genet,** v. 47, n. 5, p. 512-7, May 2015. ISSN 1546-1718. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/25848748 >.

TAKAI, H. et al. A POT1 mutation implicates defective telomere end fill-in and telomere truncations in Coats plus. **Genes Dev,** v. 30, n. 7, p. 812-26, Apr 2016. ISSN 1549-5477. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/27013236 >.

TOUZOT, F. et al. Function of Apollo (SNM1B) at telomere highlighted by a splice variant identified in a patient with Hoyeraal-Hreidarsson syndrome. **Proc Natl Acad Sci U S A,** v. 107, n. 22, p. 10097-102, Jun 2010. ISSN 1091-6490. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/20479256 >.

TOUZOT, F.; KERMASSON, L. Extended clinical and genetic spectrum associated with biallelic RTEL1 mutations. **Blood advances,** v. 1, n. 1, p. 36-46, 2016 2016. ISSN 2473-9529.

TOWNSLEY, D. M. et al. Danazol Treatment for Telomere Diseases. **N Engl J Med,** v. 375, n. 11, p. 1095-6, Sep 2016. ISSN 1533-4406. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/27626528 >.

TOWNSLEY, D. M.; DUMITRIU, B.; YOUNG, N. S. Bone marrow failure and the telomeropathies. **Blood,** v. 124, n. 18, p. 2775-83, Oct 2014. ISSN 1528-0020. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/25237198 >.

TOWNSLEY, D. M. et al. Eltrombopag Added to Standard Immunosuppression for Aplastic Anemia. **N Engl J Med,** v. 376, n. 16, p. 1540-1550, 04 2017. ISSN 1533-4406. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/28423296 >.

TOWNSLEY, D. M.; WINKLER, T. Nontransplant therapy for bone marrow failure. **Hematology Am Soc Hematol Educ Program,** v. 2016, n. 1, p. 83-89, Dec 2016. ISSN 1520-4383. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/27913466 >.

TUMMALA, H. et al. Poly(A)-specific ribonuclease deficiency impacts telomere biology and causes dyskeratosis congenita. **J Clin Invest,** v. 125, n. 5, p. 2151-60, May 2015. ISSN 1558-8238. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/25893599 >.

URINGA, E. J. et al. RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination. **Nucleic Acids Res,** v. 39, n. 5, p. 1647-55, Mar 2011. ISSN 1362-4962. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/21097466 >.

VANNIER, J. B. et al. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. **Cell,** v. 149, n. 4, p. 795-806, May 2012. ISSN 1097-4172. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/22579284 >.

VANNIER, J. B.; SAREK, G.; BOULTON, S. J. RTEL1: functions of a disease-associated helicase. **Trends Cell Biol,** v. 24, n. 7, p. 416-25, Jul 2014. ISSN 1879-3088. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/24582487 >.

VULLIAMY, T. et al. Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. **Proc Natl Acad Sci U S A,** v. 105, n. 23, p. 8073-8, Jun 2008. ISSN 1091-6490. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/18523010 >.

VULLIAMY, T. J. et al. Differences in disease severity but similar telomere lengths in genetic subgroups of patients with telomerase and shelterin mutations. **PLoS One,** v. 6, n. 9, p. e24383, 2011. ISSN 1932-6203. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/21931702 >.

VULLIAMY, T. J. et al. Mutations in the reverse transcriptase component of telomerase (TERT) in patients with bone marrow failure. **Blood Cells Mol Dis,** v. 34, n. 3, p. 257-63, 2005 May-Jun 2005. ISSN 1079-9796. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/15885610 >.

WALNE, A. J. et al. TINF2 mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. **Blood,** v. 112, n. 9, p. 3594-600, Nov 2008. ISSN 1528-0020. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/18669893 >.

WALNE, A. J. et al. Constitutional mutations in RTEL1 cause severe dyskeratosis congenita. **Am J Hum Genet,** v. 92, n. 3, p. 448-53, Mar 2013. ISSN 1537-6605. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/23453664 >.

WEI, Y. et al. Ncor2 is required for hematopoietic stem cell emergence by inhibiting Fos signaling in zebrafish. **Blood,** v. 124, n. 10, p. 1578-85, Sep 2014. ISSN 1528-0020. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/25006126 >.

WINKLER, T. et al. Defective telomere elongation and hematopoiesis from telomerase-mutant aplastic anemia iPSCs. **J Clin Invest,** v. 123, n. 5, p. 1952-63, May 2013. ISSN 1558-8238. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/23585473 >.

WU, P.; TAKAI, H.; DE LANGE, T. Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. **Cell,** v. 150, n. 1, p. 39-52, Jul 2012. ISSN 1097-4172. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/22748632 >.

WU, P. et al. Apollo contributes to G overhang maintenance and protects leading-end telomeres. **Mol Cell,** v. 39, n. 4, p. 606-17, Aug 2010. ISSN 1097-4164. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/20619712 >.

XIN, Z. T. et al. Functional characterization of natural telomerase mutations found in patients with hematologic disorders. **Blood,** v. 109, n. 2, p. 524-32, Jan 2007. ISSN 0006-4971. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/16990594 >.

YAMAGUCHI, H. et al. Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome. **Blood,** v. 102, n. 3, p. 916-8, Aug 2003. ISSN 0006-4971. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/12676774 >.

YAMAGUCHI, H. et al. Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. **N Engl J Med,** v. 352, n. 14, p. 1413-24, Apr 2005. ISSN 1533-4406. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/15814878 >.

YOSHIZATO, T. et al. Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. **N Engl J Med,** v. 373, n. 1, p. 35-47, Jul 2015. ISSN 1533-4406. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/26132940 >.

YOUNG, A. L. et al. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. **Nat Commun,** v. 7, p. 12484, Aug 2016. ISSN 2041-1723. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/27546487 >.

YOUNG, N. S. Current concepts in the pathophysiology and treatment of aplastic anemia. **Hematology Am Soc Hematol Educ Program,** v. 2013, p. 76-81, 2013. ISSN 1520-4383. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/24319166 >.

YOUNG, N. S.; CALADO, R. T.; SCHEINBERG, P. Current concepts in the pathophysiology and treatment of aplastic anemia. **Blood,** v. 108, n. 8, p. 2509-19, Oct 2006. ISSN 0006-4971. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/16778145 >.

YOUNG, N. S.; MACIEJEWSKI, J. The pathophysiology of acquired aplastic anemia. **N Engl J Med,** v. 336, n. 19, p. 1365-72, May 1997. ISSN 0028-4793. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/9134878 >.

YOUNG, N. S.; SCHEINBERG, P.; CALADO, R. T. Aplastic anemia. **Curr Opin Hematol,** v. 15, n. 3, p. 162-8, May 2008. ISSN 1065-6251. Available at: < http://www.ncbi.nlm.nih.gov/pubmed/18391779 >.

ZELLINGER, B. et al. Ku suppresses formation of telomeric circles and alternative telomere lengthening in Arabidopsis. **Mol Cell,** v. 27, n. 1, p. 163-9, Jul 2007. ISSN 1097-2765. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/17612498 >.

APPENDIX

Appendix 1. Variants classification according to a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG).

RTEL1 **variants**	**Criteria**	**Classification**
M652T (rs148080505)	-Absent in population databases based on patient ethnicity: moderate-pathogenic -Multiple lines of in silico evidence suggest no impact on gene function: supporting-benign -Highly conserved amino acid: supporting-pathogenic -Patient's phenotype: supporting-pathogenic -Functional assay (3' overhang measurement) associated to variant: supporting- pathogenic -Predicted to destabilize the helicase intra-domain interactions by 3D modeling: supporting-pathogenic	Likely pathogenic
D719N	-Highly conserved amino acid: supporting-pathogenic -Patient's phenotype and very short telomeres: supporting-pathogenic -Compound heterozygous associated with an autosomal recessive disease: moderate-pathogenic -Multiple lines of in silico evidence support a deleterious impact on gene function: supporting-pathogenic -Segregated with the disease: supporting-pathogenic	Likely pathogenic
P82L	-Functional data (3' overhang measurement) associated this variant with dysfunctional gene: supporting-pathogenic -Segregated with the disease: supporting-pathogenic -Multiple lines of in silico evidence suggest no impact on gene function: supporting-benign	Uncertain significance
P867L (rs139083375)	-Multiple lines of in silico evidence suggest no impact on gene function: supporting-benign --Co-occurred with a potential pathogenic *TERT* variant: supporting-benign	Likely benign
P871L (rs144002969)	-Detected in *trans* with the pathogenic D719N variant: moderate-pathogenic -Multiple lines of in silico evidence suggest no impact on gene function: supporting-benign -Not conserved among mammals: Strong-benign -Patient's phenotype: supporting-pathogenic	Likely benign
P884L (rs144002969)	-Not conserved among mammals: supporting-benign -Multiple lines of in silico evidence suggest no impact on gene function: supporting-benign -Functional assay (3' overhang measurement) associated with the variant: supporting-pathogenic -Absent in population databases based on patient ethnicity: moderate-pathogenic	Uncertain significance

A902G	-Absent in population databases: moderate-pathogenic -Co-occurred with a pathogenic *TERC* variant: supporting-benign -Multiple lines of in silico evidence suggest no impact on gene function: supporting-benign	Uncertain significance
G951S	-Detected in *trans* with the pathogenic F1262L variant: moderate-pathogenic -Absent in population databases: moderate-pathogenic -Multiple lines of in silico evidence support a deleterious impact on gene function: supporting-pathogenic -Segregated with the disease: supporting-pathogenic -Patient's phenotype highly specific for gene mutated: supporting-pathogenic -Highly conserved amino acid: supporting-pathogenic	Likely pathogenic
P996H	-Multiple lines of in silico evidence suggest no impact on gene function: supporting-benign -Patient’s normal TL and 3’ overhang: supporting-benign -Population frequency according to patient’s ethnicity: supporting-benign	Likely benign
F1262L	-Predicted to disrupt a domain critical for gene function: Very strong-pathogenic -Well-establish functional studies show a deleterious effect (*TRF2* downregulation): Strong-pathogenic -Compound heterozygosity associated with an autosomal recessive disease: moderate-pathogenic -Patient's short telomeres: supporting-pathogenic -Predicted to affect the RTEL1 protein partner interactions by 3D model: supporting-pathogenic	Pathogenic

The criteria used to evaluate variants’ pathogenicity were based on different types of evidence: computational and prediction data, functional and segregation data, patients’ mutational profile and clinical features, and genomic databases data. ACMG criteria also established the strength for the evidence identified as very strong, strong, moderate, and supporting for pathogenicity, or strong and supporting for a benign assertion. The type of evidence as well as its strength is described in the table.

Appendix 2. Customized SureSelectXT Target Enrichment panel with 165 genes related to bone marrow failure and hematologic

Gene	Gene function	Reported disease	Genomic position	Size (bp)	Coverage
ABL1	Oncogene	AML	chr9:133589218-133763112	7541	100
AEBP2	Epigenetic regulation	Secondary AML	chr12:19556929-19873785	7739	96.07185
ASXL1	Epigenetic regulation	MDS/AML	chr20:30946097-31027172	11026	98.16797
ATM	DNA damage and repair	MDS	chr11:108093161-108239879	28632	94.06608
ATRX	Epigenetic regulation	MDS/AML	chrX:76760306-77041805	15977	100
BCOR	Transcription related	AA/AML	chrX:39909018-40036632	9605	100
BCORL1	Transcription related	AA/AML	chrX:129115033-129192108	9696	100
BRAF	Signal transduction	AML	chr7:140419077-140624614	12946	100
BRCA2	DNA damage and repair	AML	chr13:32889561-32973859	14555	97.32051
BRCC3	Ubiquitination	MDS	chrX:154299645-154351399	4382	80.62528
BRIP1	DNA repair	Fanconi anemia	chr17:59756497-59940970	12509	88.01663
CALR	Transcription related	MPD	chr19:13049342-13055354	3375	96.68148
CBL	Ubiquitination	MDS	chr11:119076702-119178909	13053	99.41776
CBLB	Oncogene	Neoplasia	chr3:105374255-105588446	14150	98.17667
CBLC	Oncogene	neoplasia	chr19:45281076-45303953	2691	100
CCT2	Telomere biology gene	DC/AA/IPF	chr12:69979064-69995407	5929	95.32805
CCT3	Telomere biology gene	DC/AA/IPF	chr1:156278702-156337714	5817	93.79405

CCT4	Telomere biology gene	DC/AA/IPF	chr2:62095174-62115989	4314	100
CCT5	Telomere biology gene	DC/AA/IPF	chr5:10249983-10266574	6901	100
CCT6A	Telomere biology gene	DC/AA/IPF	chr7:56119273-56131732	5399	97.53658
CCT7	Telomere biology gene	DC/AA/IPF	chr2:73460498-73480200	5118	100
CCT8	Telomere biology gene	DC/AA/IPF	chr21:30428076-30446168	4796	100
CDAN1	Chromatin changes	Congenital dyserythropoietic anemia	chr15:43015707-43029467	8103	99.38294
CDH23	Cell adhesion	AML	chr10:73156641-73575754	25809	99.07009
CDKN2A	Tumor suppressor	neoplasias	chr9:21967701-21995350	5338	100
CEBPA	Transcription related	AML	chr19:33790790-33793520	2731	100
CSF3R	Receptor	AML/ Congenital neutropenia	chr1:36931594-36948965	9869	96.67646
CSMD1	Signal transduction	AML	chr8:2792825-4852544	26900	100
CTC1	Telomere biology gene	DC/AA/IPF	chr17:8128089-8151463	10303	99.12647
CTCF	Transcription related	MDS	chr16:67596260-67673138	5403	100
CUX1	Transcription related	MDS	chr7:101458909-101927300	22230	99.69411
DAXX	Transcription related	AML	chr6:33286285-33297096	4665	100
DCLRE1B (Apollo)	Telomere biology gene	DC/AA/IPF	chr1:114447713-114456758	4340	95.25346
DDX41	RNA splicing	MDS	chr5:176938528-176944520	5332	100
DHX36	Telomerase maturation	none	chr3:153990285-154042336	13263	93.09357
DIDO1	Transcriptional factor	Familial MDS	chr20:61509040-61569354	15029	100
DIS3	RNA processing	AML	chr13:73329490-73356316	9826	91.54285

DKC1	Telomere biology gene	DC/AA/IPF	chrX:153990967-154006014	6042	96.70639
DNMT3A	Epigenetic regulation	MDS/AML	chr2:25455446-25565509	10000	100
EED	Epigenetic regulation	MDS/AML	chr11:85955376-85989905	8176	97.85959
ELANE	Elastases	Congenital neutropenia	chr19:850964-856296	1627	100
ERCC4	DNA damage and repair	neoplasias	chr16:14013964-14046255	9764	91.70422
ETNK1	Phospatidyletanolamine synthesis	MDS	chr12:22777959-22843658	9300	98.29032
ETV6	Transcription related	MDS	chr12:11802738-12048386	8275	97.07553
EZH1	Epigenetic regulation	MDS/AML	chr17:40852243-40897121	9215	93.80358
EZH2	Epigenetic regulation	MDS/AML	chr7:148504414-148581491	6665	100
FANCA	DNA repair	Fanconi anemia	chr16:89803907-89883115	13691	96.16537
FANCB	DNA repair	Fanconi anemia	chrX:14861479-14891241	4215	100
FANCC	DNA repair	Fanconi anemia	chr9:97861286-98080041	10121	99.37753
FANCD2	DNA repair	Fanconi anemia	chr3:10068048-10143664	12938	89.91344
FANCE	DNA repair	Fanconi anemia	chr6:35420088-35434931	3555	96.82138
FANCF	DNA repair	Fanconi anemia	chr11:22644029-22647437	3409	97.97595
FANCG	DNA repair	Fanconi anemia	chr9:35073782-35080063	4814	100
FANCI	DNA repair	Fanconi anemia	chr15:89787130-89860542	11806	98.71252
FANCL	DNA repair	Fanconi anemia	chr2:58386328-58468565	3835	100
FANCM	DNA repair	Fanconi anemia	chr14:45605092-45670143	10969	98.75103

FBXW7	Ubiquitination	MDS	chr4:153242360-153457303	11788	98.07431
FLT3	Receptor	AML	chr13:28577361-28674779	6741	100
G6PC3	gluconeogenic and glycogenolytic pathways	Congenital neutropenia	chr17:42148048-42153762	5056	88.09335
GATA1	Transcription factor	Dysfunctional hematopoiesis	chrX:48644912-48652768	2100	100
GATA 2	Transcription factor	GATA2 deficiency			
GFI1	Transcription related	Congenital neutropenia	chr1:92940268-92952483	4302	100
GNAS	Signal transduction	MDS	chr20:57414723-57486300	15498	98.90308
GPRC5A	Signal transduction	MDS	chr12:13043666-13070921	8840	93.84615
HAX1	Signal transduction	Congenital neutropenia	chr1:154244937-154248405	2550	100
HRAS	Oncogene/DNA repair	neoplasias	chr11:532192-537337	2783	100
HSP90AA 1	Signal transduction	none	chr14:102547025-102606136	6347	97.95179
IDH1	Epigenetic regulation	MDS/AML	chr2:209100901-209130848	6086	95.59645
IDH2	Epigenetic regulation	MDS/AML	chr15:90626227-90645836	4040	97.45049
IKZF1	Regulation of telomere maintenance	ALL	chr7:50343629-50472849	11102	100
IRF1	Transcription related	MDS	chr5:131817251-131826540	8199	89.65728
JAK1	Signal transduction	MPD	chr1:65298856-65432237	9023	100
JAK2	Signal transduction	MPD	chr9:4984983-5128233	8739	100
JAK3	Signal transduction	MPD	chr19:17935539-17958930	9166	93.12677
JARID2	Epigenetic regulation	MDS/AML	chr6:15246156-15522323	9080	100

KDM6A	Epigenetic regulation	MDS/AML	chrX:44732371-44971907	10127	100
KIT	Receptor	AML	chr4:55524035-55606931	7637	100
KMT2A (MLL)	Transcription related	AML	chr11:118307155-118397589	25495	98.46637
KRAS	Oncogene	neoplasias	chr12:25357673-25403920	8002	98.52537
LAMB4	Cell adhesion	MDS	chr7:107663943-107770851	10716	98.26428
LUC7L2	RNA splicing	MDS	chr7:139025055-139108253	7934	97.71868
MAP3K4	Signal transduction	AML	chr6:161412709-161538467	10522	98.11823
MPL	Receptor	AA/MDS	chr1:43803425-43820185	5023	88.57257
MRE11A	DNA damage and repair	Ataxia-telangiectasia-like disorder	chr11:94150419-94227124	8164	95.23518
MYC	Oncogene	AML	chr8:128747630-128753730	3301	100
MYD88	Signal transduction	Cancer differentiation	chr3:38179919-38184563	4389	100
NAF1	Telomerase biogenesis	none	chr4:164031175-164088123	6108	97.7243
NBN (NBS1)	DNA damage and repair	AA	chr8:90945514-91015506	8304	95.42389
NCOR2	Transcription related	MDS	chr12:124808907-125052185	19756	98.92184
NF1	Signal transduction	MDS	chr17:29421895-29709184	32872	98.00134
NHP2	Telomere biology gene	DC	chr5:177576411-177581018	1888	91.31356
NOP10	Telomere biology gene	DC	chr15:34633864-34635428	754	100
NOTCH1	Signal transduction	CLL	chr9:139388846-139440364	13217	100
NPM1	Transcription related	AML	chr5:170814070-170838191	4928	97.1388

NRAS	Oncogene	neoplasias	chr1:115247035-115259565	5154	100
OBFC1	Telomere maintenance	none	chr10:105637268-105678095	7781	100
PALB2 (FANCN)	DNA repair	Fanconi anemia	chr16:23614433-23652728	6292	100
PARN	Telomere biology gene	DC/AA/IPF	chr16:14529507-14726635	7464	94.3194
PCNA	DNA damage and repair	none	chr20:5095549-5107322	2148	100
PDGFRA	Signal transduction	CML	chr4:55095214-55164464	12623	96.72028
PEG3	Apoptosis	Acquired AA	chr19:57321395-57352146	10484	100
PHF6	Transcription related	MDS	chrX:133507233-133562872	9017	81.80105
PIF1	Telomerase regulator	none	chr15:65107779-65117917	5712	92.33193
PIGA	GPI anchor	Acquired AA	chrX:15337523-15353726	4346	100
PML	Tumor suppressor	AML	chr15:74286964-74340205	13593	95.83609
POT1	Telomere biology gene	DC/AA/IPF	chr7:124462390-124570087	8430	95.05338
PRF1	Perforin	Acquired AA	chr10:72357054-72362581	2834	99.22371
PRPF8	RNA splicing	MDS	chr17:1553873-1588226	15283	94.07185
PTEN	Tumor suppressor	neoplasias	chr10:89622820-89731737	11307	98.08084
PTGES3 (p23)	Telomerase related	none	chr12:57057075-57082242	4299	94.7197
PTPN11	Signal transduction	MDS	chr12:112856105-112947767	8796	94.88404
RAD21	Cohesin	AML	chr8:117858123-117887155	7575	100
RAD51C	DNA repair	Fanconi anemia	chr17:56769884-56811753	4495	95.10568
RB1	Oncogene	neoplasms	chr13:48877833-49056172	8946	99.59759

RBBP4	Epigenetic regulation	MDS/AML	chr1:33116693-33151862	10835	92.49654
RBBP7	Epigenetic regulation	MDS/AML	chrX:16857356-16888587	8492	88.70702
RIT1	Signal transduction	MDS	chr1:155867549-155881245	4619	100
RPL11	Ribosomal related gene	DBA	chr1:24018219-24022965	2592	100
RPL15	Ribosomal related gene	DBA	chr3:23957986-23965237	6732	83.70469
RPL35A	Ribosomal related gene	DBA	chr3:197676808-197683531	3677	84.14468
RPL5	Ribosomal related gene	DBA	chr1:93297532-93307531	3700	99.18919
RPS10	Ribosomal related gene	DBA	chr6:34385181-34393952	2353	100
RPS17	Ribosomal related gene	DBA	chr15:82821108-83209345	5404	98.29756
RPS19	Ribosomal related gene	DBA	chr19:42363938-42377044	4605	96.19978
RPS24	Ribosomal related gene	DBA	chr10:79793468-79816621	5779	96.46998
RPS26	Ribosomal related gene	DBA	chr12:56435587-56438166	1900	100
RPS7	Ribosomal related gene	DBA	chr2:3622745-3628559	5783	98.30538
RTEL1	Telomere biology gene	DC/AA/IPF	chr20:62289113-62330087	13773	100
RUNX1	Transcription related	MDS/AML	chr21:36160048-37377015	18566	95.68566
RUVBL1	TERC maturation	None	chr3:127783571-127872807	8167	98.76331
RUVBL2	TERC maturation	none	chr19:49496655-49519302	5561	95.34257
SBDS	Ribosomal related gene	SDS	chr7:66452614-66460685	2471	100
SETBP1	Oncogene	MDS	chr18:42260088-42648525	12061	99.97513
SF3B1	RNA splicing	MDS	chr2:198254458-198299865	11742	98.9099
SH2B3	Signal transduction	AA/MDS	chr12:111843702-111889477	6660	99.12913

SHQ1	Telomerase biogenesis	none	chr3:72798378-72911115	7441	94.18089
SLX4 (FANCP)	DNA repair	Fanconi anemia	chr16:3631132-3661649	9991	95.85627
SMC1A	Cohesin	AML	chrX:53401020-53449727	13275	99.54049
SMC3	Cohesin	AML	chr10:112327399-112364444	7256	100
SRP72	Signal transduction	AA/MDS	chr4:57333031-57369897	6563	100
SRSF2	RNA splicing	MDS	chr17:74730147-74733543	3254	100
STAG2	Cohesin	AML	chrX:123094012-123556564	12578	96.9709
STAT3	Signal transduction	AA/MDS	chr17:40465292-40540636	10818	96.45036
SUZ12	Epigenetic regulation	MDS/AML	chr17:30263987-30328114	9379	96.67342
TCP1	TriC related - telomeres	None	chr6:160199480-160210831	5917	100
TEN1	Telomere maintenance	None	chr17:73975248-73996717	1906	98.26863
TERC	Telomere biology gene	DC/AA/IPF	chr3:169482258-169482898	641	100
TERF1	Telomere biology gene	DC/AA/IPF	chr8:73921047-73960407	9933	77.69052
TERF2	Telomere biology gene	DC/AA/IPF	chr16:69389414-69442524	6743	96.47041
TERT	Telomere biology gene	DC/AA/IPF	chr5:1253212-1295234	6216	100
TET2	Epigenetic regulation	MDS/AML	chr4:106066982-106201023	17736	91.68358
TGS1	Telomerase maturation	none	chr8:56685651-56738057	5169	98.60708
THPO	Signal transduction	AA/Thrombocytopenia	chr3:184089673-184097526	2946	100
TINF2	Telomere biology gene	DC/AA/IPF	chr14:24708799-24711930	3075	98.60162
TNKS	Signal transduction	neoplasias	chr8:9413374-9639906	17974	99.02081

TP53	Oncogene	MDS/AML	chr17:7565047-7590918	5309	88.15219
TPP1	Telomere biology gene	DC	chr11:6633947-6640742	5706	100
U2AF1	RNA splicing	MDS	chr21:44513016-44527747	11503	98.04399
U2AF2	RNA splicing	MDS	chr19:56165366-56186132	5525	98.53394
UMODL1	Cell adhesion	AML	chr21:43483018-43563613	10278	99.69838
WAS	Signal transduction	Thrombocytopenia	chrX:48534935-48549868	3754	98.90784
WRAP53	Telomere biology gene	DC/AA/IPF	chr17:7589339-7606870	5864	100
WT1	Transcription related	AML	chr11:32409271-32457226	5417	100
ZRSR2	RNA splicing	MDS	chrX:15808524-15841433	3298	91.14615
ZSWIM4	Cell adhesion	AML	chr19:13906224-13943094	6098	100

Abbreviations as followed: bp, base pairs; AML, acute myeloid leukemia; MDS, myelodysplastic syndrome; DC, dyskeratosis congenital; AA, aplastic anemia; IPF, idiopathic pulmonary fibrosis; ALL, acute lymphoblastic leukemia; MPD, Myeloproliferative diseases; CLL, Chronic lymphocytic leukemia; DBA, Diamond–Blackfan anemia;

Appendix 3. The genes included in the targeted NGS panel used by the NIH cohort.

NIH cohort:

Genes related to Fanconi anemia*

BRCA2 (FANCD1), BRIP1 (FANCJ), ERCC4 (FANCQ), FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, PALB2 (FANCN), RAD51C (FANCO), SLX4 (FANCP), XRCC2

Genes related to ribosomal diseases*

RPL5, RPL11, RPL35A, RPS7, RPS10, RPS19, RPS24, RPS26

Genes related to telomere diseases*

CTC1, DKC1, NHP2 (NOLA2), NOP10 (NOLA3), RTEL1, TERC, TERT, TINF2, USB1, WRAP53

Genes previously reported associated to myeloid malignancies*

RUNX1, SRP72

Severe congenital neutropenia (SCN)*

CSF3R, ELANE, G6PC3, GFI1, HAX1, VPS45, WAS.

Others*

GATA2, MPL, RBM8A, SBF2, SBDS.

* Inherited Bone Marrow Failure Sequencing Panel generated by the University of Chicago.

Appendix 4 - Parecer consubstanciado da aprovação do projeto de pesquisa pela Comissão de Ética em pesquisa (CEP) do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto/USP.

HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO

FMRP-USP RIBEIRÃO PRETO

Ribeirão Preto, 19 de agosto de 2015

Ofício nº 2991/2015
CEP/MGV

Prezados Senhores,

O trabalho intitulado **"IDENTIFICAÇÃO DE MODULADORES GENÉTICOS DA RESPOSTA À IMUNOSSUPRESSÃO NA ANEMIA APLÁSTICA ADQUIRIDA POR SEQUENCIAMENTO DE NOVA GERAÇÃO" – versão 2, de 12/05/2015,** foi analisado pelo Comitê de Ética em Pesquisa, em sua 413ª Reunião Ordinária realizada em 17/08/2015 e enquadrado na categoria: **APROVADO,** bem como o **Termo de Consentimento Livre e Esclarecido – versão 2, de 12/05/2015, o Biorrepositório "SANGUE DE PACIENTES COM ANEMIA APLÁSTICA ADQUIRIDA" e o Termo de Guarda de Material Biológico,** de acordo com o Processo HCRP nº 203/2015.

De acordo com Carta Circular nº 003/2011/CONEP/CNS, datada de 21/03/2011, o sujeito de pesquisa ou seu representante, quando for o caso, deverá rubricar todas as folhas do Termo de Consentimento Livre e Esclarecido – TCLE – apondo sua assinatura na última do referido Termo; o pesquisador responsável deverá da mesma forma, rubricar todas as folhas do Termo de Consentimento Livre e Esclarecido – TCLE – apondo sua assinatura na última página do referido Termo.

Este Comitê segue integralmente a Conferência Internacional de Harmonização de Boas Práticas Clínicas (IGH-GCP), bem como a Resolução nº 466/12 CNS/MS.

Lembramos que devem ser apresentados a este CEP, o Relatório Parcial e o Relatório Final da pesquisa.

Atenciosamente.

DRª. MARCIA GUIMARÃES VILLANOVA
Coordenadora do Comitê de Ética em
Pesquisa do HCRP e da FMRP-USP

Ilustríssimos Senhores
FERNANDA GUTIERREZ RODRIGUES
PROF.DR. RODRIGO DO TOCANTINS CALADO DE SALOMA RODRIGUES
Depto. de Clínica Médica