Gain-of-Function Mutations in RPA1 Cause a Syndrome with Short Telomeres and Somatic Genetic Rescue

Richa Sharma, Sushree Sahoo, Masayoshi Honda, Sophie Granger, Charnise Goodings, Louis Sanchez, Axel Künstner, Hauke Busch, Fabian Beier, Shondra Pruett-Miller, et al.

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Gain-of-Function Mutations in RPA1 Cause a Syndrome with Short Telomeres and Somatic Genetic Rescue

Tracking no: BLD-2021-011980R3

Richa Sharma (St Jude Children's Research Hospital, United States) Sushree Sahoo (St Jude Children's Research Hospital, United States) Masayoshi Honda (University of Iowa, United States) Sophie Granger (University of Iowa, United States) Charnise Goodings (St. Jude Children's Research Hospital, United States) Louis Sanchez (Sorbonne Université, Germany) Axel Künstner (University of Lübeck, Germany) Hauke Busch (University of Luebeck, Germany) Fabian Beier (University Hospital Aachen, Germany) Shondra Pruett-Miller (St. Jude Children's Research Hospital, United States) marcus Valentine (St Jude Children's Research Hospital, United States) Alfonso Fernandez (St Jude Children's Research Hospital, United States) Ti-Cheng Chang (St. Jude Children's Research Hospital, United States) Vincent Géli (U1068 Inserm, UMR7258 CNRS, Aix-Marseille University, Institut Paoli-Calmettes, France) Dmitri Churikov (Marseille Cancer Research Center, France) Sandrine Hirschi (Strasbourg University Hospital, University of Strasbourg, France) Victor Pastor (St Jude Children's Research Hospital, United States) Melanie Boerries (German Cancer Consortium (DKTK), Germany) Melchior Lauten (University Hospital Schleswig-Holstein, Campus Lübeck, Germany) Charikleia Kelaidi (Aghia Sophia Children's Hospital, Greece) Undiagnosed Network (,) Megan Cooper (Washington University, United States) Sarah Nicholas (Baylor College of Medicine, United States) Jill Rosenfeld (Baylor College of Medicine, United States) Sophia Polychronopoulou (Aghia Sophia Children's Hospital, Greece) Caroline Kannengiesser (APHP, Hopital Xavier Bichat, Service de Génétique, France) Carole Saintome (Muséum National d'Histoire Naturelle, France) Charlotte Niemeyer (German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Germany) Patrick Revy (Imagine Institute for Genetic Diseases, France) Marc Wold (University of Iowa, United States) Maria Spies (University of Iowa, United States) Miriam Erlacher (German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ),) Stéphane Coulon (Marseille Cancer Research Centre, U1068 INSERM, UMR7258 CNRS, UM105 Aix-Marseille University, Institut Paoli-Calmettes,, France) Marcin Wlodarski (University of Freiburg, Germany)

Abstract:
Human telomere biology disorders (TBD)/short telomere syndromes (STS) are heterogeneous disorders caused by inherited loss-of-function mutations in telomere-associated genes. Here, we identify three germline heterozygous missense variants in *RPA1* gene in four unrelated probands presenting with short telomeres and varying clinical features of TBD/STS including bone marrow failure, myelodysplastic syndrome, T- and B-cell lymphopenia, pulmonary fibrosis, or skin manifestations. All variants cluster to DNA binding domain A of RPA1 protein. RPA1 is a single-strand DNA-binding protein required for DNA replication and repair and involved in telomere maintenance. We showed that RPA1^{E240K} and RPA1^{V227A} proteins exhibit increased binding to single-strand and telomeric DNA, implying a gain in DNA-binding function while RPA1^{T270A} has binding properties similar to wild type protein. To study the mutational effect in a cellular system, we used CRISPR/Cas9 to knock-in the *RPA1*^{E240K} mutation into healthy inducible pluripotent stem cells. This resulted in severe telomere shortening and impaired hematopoietic differentiation. Furthermore, in patient with *RPA1*^{E240K}, we discovered somatic genetic rescue (SGR) in hematopoietic cells due to an acquired truncating *cis RPA1* mutation or a uniparental isodisomy 17p with loss of mutant allele, coinciding with stabilized blood counts. Using single-cell sequencing, the two SGR events were proven to be independently acquired in hematopoietic stem cells. In summary, we describe the first human disease caused by germline *RPA1* variants in individuals with TBD/STS.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: RS, SSS, MWW, CS, SC, MSW, MS conceived and designed the experiments. RS, SSS, AK, HB, TCC, MB, VPL, JAR, CK, MWW performed genomic data analysis. RS, MH, SLG, CG, LS, FB, MBV, SMPM, AGF, DC, VG, CS, PR, MSW, MS, and SC performed and/or interpreted functional experiments. FB, SH, ML, CK, MC, SN, JAR, SP, CMN, ME, RS, MWW were involved in patient care, collecting clinical data and clinical testing. MWW, SC, MSW, MS supervised the experiments, and MWW oversaw study design. All authors contributed to the manuscript and approved of the final version.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: We are currently in the process of submitting all sequencing files to EGA archive (EGA box 1017)

Clinical trial registration information (if any):

Gain-of-Function Mutations in RPA1 Cause a Syndrome with Short Telomeres and Somatic Genetic Rescue

Richa Sharma[1,#], Sushree S. Sahoo[1,#], Masayoshi Honda[2], Sophie L. Granger[2], Charnise Goodings[1], Louis Sanchez[3,4], Axel Künstner[5], Hauke Busch[5], Fabian Beier[6], Shondra M. Pruett-Miller[7], Marcus B. Valentine[8], Alfonso G. Fernandez[1], Ti-Cheng Chang[9], Vincent Géli[10], Dmitri Churikov[10], Sandrine Hirschi[11], Victor B. Pastor[9], Melanie Boerries[12,13], Melchior Lauten[14], Charikleia Kelaidi[15], Undiagnosed Disease Network[&], Megan A. Cooper[16], Sarah Nicholas[17], Jill A. Rosenfeld[18], Sophia Polychronopoulou[15], Caroline Kannengiesser[19], Carole Saintomé[3,4], Charlotte M. Niemeyer[13,20], Patrick Revy[21], Marc S. Wold[2], Maria Spies[2], Miriam Erlacher[13,20], Stéphane Coulon[10], Marcin W. Wlodarski[1,20*]

[1]Department of Hematology, St. Jude Children´s Research Hospital, Memphis, TN, USA
[2]Department of Biochemistry, Carver College of Medicine, University of Iowa, Iowa City, IA, USA
[3]Structure et Instabilité des Génomes, Muséum National d'Histoire Naturelle, CNRS UMR 7196, INSERM U1154, 43 rue Cuvier, F-75005, Paris, France
[4]Sorbonne Université, UFR927, F-75005, Paris, France
[5]Lübeck Institute of Experimental Dermatology and Institute of Cardiogenetics, University of Lübeck, Lübeck, Germany
[6]Department of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation, Medical Faculty, RWTH Aachen University, Aachen, Germany
[7]Depatment of Cell and Molecular Biology, St. Jude Children´s Research Hospital, Memphis, TN, USA
[8]Cytogenetics Core Facility, St. Jude Children's Research Hospital, Memphis, TN
[9]Center for Applied Bioinformatics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
[10]Marseille Cancer Research Centre, U1068 INSERM, UMR7258 CNRS, UM105 Aix-Marseille University, Institut Paoli-Calmettes, Equipe labellisée par la Ligue Nationale contre le Cancer, Marseille, F-13009, France
[11]Department of Respiratory Medicine and Rare Pulmonary Diseases, Strasbourg University Hospital, Strasbourg, France
[12]Institute of Medical Bioinformatics and Systems Medicine, Medical Center-University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[13]German Cancer Consortium (DKTK), Freiburg, Germany and German Cancer Research Center (DKFZ), Heidelberg
[14]University Hospital Schleswig-Holstein, Department of Pediatrics, University of Lübeck, Germany
[15]Department of Pediatric Hematology/Oncology, Aghia Sophia Children's Hospital, Athens, Greece
[16]Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA
[17]Department of Allergy and Immunology, Baylor College of Medicine, Houston, TX, USA
[18]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
[19]Department of Genetics, Bichat Hospital, Assistance Publique - Hôpitaux de Paris, Paris University, INSERM U1152, Paris, France
[20]Division of Pediatric Hematology and Oncology, Department of Pediatrics and Adolescent Medicine, Medical Center, University of Freiburg, Freiburg, Germany
[21]Université de Paris, Imagine Institute, Laboratory of Genome Dynamics in the Immune System, Laboratoire labellisé Ligue, INSERM UMR 1163, F-75015, Paris, France
# Equal contribution
& Full contributor list of the Undiagnosed Disease Network in the supplement

*Corresponding author:
Marcin W. Wlodarski, MD, PhD
St. Jude Children's Research Hospital
262 Danny Thomas Place, MS 341
Memphis, TN 38105
Phone: 901-595-2484, Fax: 901-595-2176
E-mail: marcin.wlodarski@stjude.org

Running title: RPA1 in human disease
Article type: Regular article
Abstract: 222, Main text: 3993 words

Keywords: RPA1, telomeres, DNA binding, telomere biology disorder, short telomere syndrome

Key points:

- Germline *RPA1* gain-of-function missense mutations result in a telomere biology disorder phenotype
- Somatic rescue events arise in hematopoiesis secondary to germline *RPA1* mutation

ABSTRACT

Human telomere biology disorders (TBD)/short telomere syndromes (STS) are heterogeneous disorders caused by inherited loss-of-function mutations in telomere-associated genes. Here, we identify three germline heterozygous missense variants in *RPA1* gene in four unrelated probands presenting with short telomeres and varying clinical features of TBD/STS including bone marrow failure, myelodysplastic syndrome, T- and B-cell lymphopenia, pulmonary fibrosis, or skin manifestations. All variants cluster to DNA binding domain A of RPA1 protein. RPA1 is a single-strand DNA-binding protein required for DNA replication and repair and involved in telomere maintenance. We showed that RPA1^{E240K} and RPA1^{V227A} proteins exhibit increased binding to single-strand and telomeric DNA, implying a gain in DNA-binding function while RPA1^{T270A} has binding properties similar to wild type protein. To study the mutational effect in a cellular system, we used CRISPR/Cas9 to knock-in the *RPA1*^{E240K} mutation into healthy inducible pluripotent stem cells. This resulted in severe telomere shortening and impaired hematopoietic differentiation. Furthermore, in patient with *RPA1*^{E240K}, we discovered somatic genetic rescue (SGR) in hematopoietic cells due to an acquired truncating *cis RPA1* mutation or a uniparental isodisomy 17p with loss of mutant allele, coinciding with stabilized blood counts. Using single-cell sequencing, the two SGR events were proven to be independently acquired in hematopoietic stem cells. In summary, we describe the first human disease caused by germline *RPA1* variants in individuals with TBD/STS.

INTRODUCTION

Telomeres are complex structures made up of repetitive DNA sequences associated with specialized proteins found at natural ends of linear chromosomes in all mammals. The hallmark function of telomeres is to protect chromosomal ends from degradation and inappropriate recombination, and activation of the DNA damage response[1]. In the absence of telomere-associated proteins, chromosomal ends undergo premature attrition with pathological consequence of telomere biology disorders (TBD), also referred to as short telomere syndromes (STS) or dyskeratosis congenita. TBD/STS are hereditary disorders that manifest on a wide phenotypic and age spectrum because of genetic heterogeneity and variable expressivity. TBD/STS associated features include bone marrow failure (BMF), pulmonary and liver fibrosis, mucocutaneous fragility, and predisposition to myelodysplastic syndromes (MDS) and cancer[2-5]. Monoallelic or biallelic inactivation in 14 genes that code for telomerase holoenzyme, shelterin complex, telomere capping machinery, and accessory telomere processes, have been identified to cause TBD/STS[2,6]. However, an estimated 30% of individuals with TBD/STS do not have a genetic resolve, which obscures timely diagnosis and clinical management[6].

Replication Protein A (RPA) is a ubiquitous, single-strand (ss) DNA binding protein that exists as a heterotrimer complex composed of RPA1, RPA2 and RPA3 in eukaryotes[7]. RPA is essential for DNA replication and repair. It binds all ssDNA sequences with high affinity to provide nuclease protection, prevent hairpin formation and recruit numerous proteins to facilitate ssDNA processing and repair. Importantly, RPA is involved in several DNA damage signaling and repair pathways including nucleotide excision repair, base excision repair, mismatch repair, and double strand break repair[8-11]. RPA1, the largest subunit of RPA, harbors four DNA binding domains (DBD, -F, -A, -B, -C) of which DBD-A and DBD-B are important for high affinity ssDNA binding and facilitating protein-protein interactions[12,13]. Despite no clear association with human disease, the ubiquitous involvement of RPA1 in genome integrity and the demonstrated overexpression in some cancers[14-17] has driven research efforts to understand its functions. Thus far, only loss-of-function (LOF) *RPA1* alleles have been identified through mutagenic screens in eukaryotic systems[18-20]. Disrupted DNA repair and chromosomal rearrangements were shown in yeast and human cell lines[18-20] while genomic instability and tumorigenesis were demonstrated in mice[21,22]. Studies in yeast revealed that RPA complex binds to telomeric regions during S-phase and unfolds G-quadruplex (G4) secondary DNA structures enriched at telomeres[23-28]. Furthermore, *RPA1* mutants were associated with telomere shortening in yeast models and cell lines[24,25,27,29-32]. However, the mechanism of how *RPA1* genetic disruption affects telomere length is not well understood.

RPA1 mutations have not yet been reported in human syndromes. Here, we describe three germline heterozygous missense variants in *RPA1* in four unrelated individuals with short telomeres and phenotypes seen in TBD/STS, including hematologic, pulmonary, and skin manifestations. Biochemical studies revealed that these mutants have core DNA-binding domains with elevated affinity, pointing to a unique gain-of-function (GOF) effect for at least two of the mutants studied, while cellular studies show defective hematopoiesis with telomere shortening in *RPA1*-mutated iPSCs. Finally, we found independent acquired somatic rescue events in one patient, resulting in inactivation of the germline *RPA1* mutation.

MATERIAL AND METHODS

Patient cohort

This study was approved by the institutional review boards of the respective institutions (St. Jude Children's Hospital, INSIGHT-HD; University of Freiburg, CPMP/ICH/135/95 and 430/16; Baylor College of Medicine, H-34433; HUPNVS, Paris 7 University, AP-HP, IRB 00006477). Written informed consent was obtained from patients or guardians. Patient 1 (P1) and patient 2 (P2) were enrolled in the observational study of the European Working Group for MDS in Childhood (EWOG-MDS, (NCT00662090). Patient (P3) had an institutional diagnosis of idiopathic pulmonary fibrosis (IPF) and was enrolled at Paris 7 University, AP-HP, IRB 00006477. Patient 4 (P4) was enrolled in Undiagnosed Disease Network (UDN) (NIH IRB 15HG0130). Peripheral blood, bone marrow, fibroblasts, and hair follicles were collected from patients and family members when available.

Whole exome sequencing

Whole exome sequencing (WES) was performed on the genomic DNA of bone marrow and peripheral blood cells of P1-P4 and family members (P1: parents and unaffected brother, P2: parents and unaffected sister, P4: parents). Standard WES sequencing and analytic approaches were applied. In detail, TruSeq DNA Exome kit (Illumina, cat no. 20020615), SureSelect V5 (Agilent, cat no. 5190-6209), and Human Comprehensive Exome (TWIST Biosciences) were used for enrichment and library preparation according to the manufacturers' instructions. The generated libraries were sequenced on the Illumina HiSeq 2500 with 150-bp paired-end reads and average 50-100x coverage. After demultiplexing, the FASTQ files were trimmed for adapter sequences and mapped either to GRCh38/hg38 (P1_quartet) or GRCh37/hg19 (P2_quartet, P3, P4_trio) using BWA MEM (v0.7.15). The mapping data from P1_trio, P2_quartet, and P3 were post-processed according to GATK best practices (GATK v4.1.7; SAMtools and PICARD v 2.23.0), and variant calling (single nucleotide variants and short insertions/deletions) was performed by applying GATK Haplotypecaller v4 for P1_trio and P3, and DeepVariant (v1.1.0; ES mode) for P2_quartet. The called variants were then annotated using ANNOVAR for P1_trio and variant effector predictor (VEP, v100) for P2_quartet. While for P3, the variants annotation and filtering process was performed using the Polyweb software interface of the Bioinformatic department of Imagine Institute, Paris. Finally, for P4_trio, the mapping, processing, variant calling, annotation, and filtering was achieved using the Baylor College of Medicine institutional pipeline[33]. After mapping, variant calling and annotation, we used a single stringent filtering approach to establish a list of rare candidate variants fulfilling the following criteria: 1) minor allelic frequency (MAF)

below 0.01% in gnomAD population database, 2) minimum 6 altered forward and reverse reads, 3) exonic non-synonymous variant effect, and 4) CADD score greater than 20. Altogether, this resulted in the discovery of *RPA1* variants reported in this study.

Single-cell DNA and protein sequencing

Single-cell DNA sequencing (scDNAseq) of bone marrow (BM) cells was performed using a custom targeted panel on the Tapestri platform (MissionBio). Briefly, a set of amplicons including RPA1 variants found in P1 (germline c.718G>A, chr17:1782314:G>A and somatic c.1735G>T, RPA1:chr17:1798378:G>T) were amplified as previously reported[34], and outlined in detail in the supplement. In addition, we used oligonucleotide-conjugated antibodies targeting CD3, CD11b, CD19, CD34, CD38, CD45RA, and CD90 cell surface proteins.

Data availability

Python-based Mosaic package for analysis of scDNA sequencing is available at GitHub (https://github.com/MissionBio/mosaic). Raw datasets of whole exome sequencing and scDNAseq have been deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/) hosted by the European Bioinformatics Institute under accession numbers EGAS00001005761 (WES) and EGAS00001005762 (single cell DNA sequencing and cell-surface protein analysis).

Additional methods are outlined in detail in the supplement: telomere length assessment (Q-FISH, Southern blot, Flow-FISH, TeSLA); biochemical studies of RPA variants (expression, purification, förster resonance energy transfer (FRET)); iPSC model (CRISPR/Cas9 targeted mutagenesis, quality assessment, hematopoietic differentiation, cytospins, protein expression, irradiation and assessment of DNA damage, flow cytometry); genomics (deep sequencing, bone marrow derived single colony sequencing, RNA sequencing, *RPA1* allelic quantification, digital droplet PCR, SNP arrays, haplotype phasing and quantification of uniparental isodisomy, Sanger sequencing); and statistics.

RESULTS

Heterogenous clinical manifestations of TBD/STS unified by short telomeres in the study cohort

In the process of identifying new genetic causes of Mendelian disorders, we compiled an international cohort of four patients with phenotype consistent with TBD/STS and yet unidentified genetic cause (Table 1, Fig. 1A). All patients' medical records and evaluations underwent central review and had negative workup for inherited BMF syndromes and TBD/STS associated genes. Unbiased WES found three germline heterozygous missense variant alleles in *RPA1* (NM_002945.5) as the most plausible candidates in our patient cohort (Table 1, Fig. 1A). *De novo RPA1* c.718G>A, p.E240K and c.808A>G, p.T270A occur in P1 and P4, respectively and are previously unreported in population or disease databases. *RPA1* c.680T>C, p.V227A variant found in P2 and P3 occurs at ultra-low frequency in control populations (Table 1). P2 father and sister are carriers without history of hematologic manifestations (clinical workup declined) and P3 family members were not available for evaluation. We did not identify other *de novo* variants in genes involved in hematopoiesis, telomere homeostasis, or cancer that would be compatible with the observed phenotypes. All *de novo* variants in P1, P2, and P4 are shown in Supplemental Table 1. *RPA1* is located on chromosome 17p13.3 and has 17 coding exons coding for a 70kDa protein that is expressed in all tissues (Fig. 1B, top panel). The three *RPA1* variants cluster to DNA binding domain A (DBD-A) of RPA1 protein (Fig. 1B, bottom panel) and are near ssDNA binding grooves (Supplemental Fig. 1A). These variants affect semi to highly conserved nucleotides (Supplemental Fig. 1B) with resulting amino acid substitutions predicted as likely deleterious (Table 1).

Our patient cohort had a broad spectrum of clinical presentation and age of onset. P1 presented at age 10 years with pancytopenia, hypoplastic BM (Fig. 1C) and history of congenital eye anomaly requiring enucleation as an infant (Table 1). She further developed classic dyskeratosis congenita associated mucocutaneous triad during adolescence (Fig. 1C). However, her clinical course was atypical due to stabilization of blood counts and mucocutaneous features (Fig. 1C) without intervention over 18 years. P2, with pre-existing facial dysmorphisms (Table 1), was diagnosed at 13 years of age with MDS with excess blasts (Fig. 1C) and somatic *NRAS* c.35G>A, p.G12D mutation at 37% allelic frequency. Mild restriction was noted on pulmonary function testing prior to receiving myeloablative allogeneic hematopoietic stem cell transplantation (HSCT). Short-term complications of HSCT included severe skin and liver graft versus host disease (GVHD) with development of necrotizing pneumonitis with extensive cavitation (Fig. 1C) secondary to multifactorial cause including pulmonary GVHD, infections (nocardia pneumonitis, pulmonary aspergillosis) and pulmonary fibrosis at 8 months post-HSCT. Patient passed away 11 months post-HSCT of multi-organ failure in setting of subarachnoid hemorrhage and intractable

chronic GVHD. P3 had early hair graying and adult-onset IPF that progressed in her 5th decade of life (Table 1, Fig. 1C). Two siblings of P3 were also diagnosed with IPF but genetic studies to segregate *RPA1* mutation were not possible due to death caused by IPF-related complications (Supplemental Fig. 2). P4 presented at birth with prematurity, failure to thrive and low T-cell receptor excision circles, triggering further workup that was significant for T- and B-cell lymphopenia and severe hypogammaglobulinemia requiring chronic replacement therapy (Supplemental Table 2). Due to clinical features consistent with TBD/STS disease spectrum, we assessed telomere length in all patients using different standard approaches depending on specimen availability. P1 and P4 had decreased telomere length in peripheral blood lymphocytes compared to age-matched controls (Fig. 1D) using flow FISH[35]. Specifically, P1 had telomere length < 1st percentile measured at two time points three years apart while P4 telomere length decreased to <1st percentile from 5th percentile over 1.5 years. Telomere restriction fragment (TRF) analysis on DNA from peripheral blood of P2 and P3 revealed short telomeres in P2 compared to family members, as well as, in P3 compared to age-matched healthy control (Fig. 1E). To gain further resolution, TeSLA assay was performed in blood DNA, showing higher proportion of telomeres less than 1 kilobase (kb) in P2 and P3 (Supplemental Fig. 3A-B), consistent with increased frequency of very short telomeres.

Germline *RPA1* variants exhibit increased binding to ssDNA and telomeric DNA

To assess the effects of human *RPA1* variants on RPA heterotrimer formation and function (Fig. 2A), we expressed and purified RPA heterotrimers with wild type RPA1 (RPA^{WT}) and mutant V227A (RPA^{V227A}), E240K (RPA^{E240K}) and T270A (RPA^{T270A}) proteins (Supplemental Fig. 4A). All proteins did form heterotrimers and bound ssDNA with high affinity in a FRET based DNA-binding assay using Cy3/Cy5-labeled dT30 ssDNA (Fig. 2B). All three mutant proteins appear to form an altered complex with ssDNA. Each saturates dT30 ssDNA at a lower stoichiometry than wild type: the binding ratio was 2.2 for RPA^{WT}, 1.3 for RPA^{V227A}, 1.1 for RPA^{E240K}, and 1.6 for RPA^{T270A} (Fig. 2B, Supplemental Table 3). We next examined RPA binding to dT15, which only interacts with DBD-A and DBD-B of RPA1. Compared to RPA^{WT}, RPA^{V227A} and RPA^{E240K} displayed a two-fold and eight-fold higher affinity, respectively, while RPA^{T270A} had a marginally higher (1.3-fold) binding (Fig. 2C, Supplemental Table 3). When assessing for binding to 15-mer human telomeric sequence $(TTAGGG)_2TTA$, RPA^{V227A} and RPA^{E240K} had increased binding affinity while RPA^{T270A} had no effect compared to RPA^{WT} (Fig. 2D, Supplemental Table 3). Next, we examined the kinetics of the telomeric G-quadruplexes (h-telG4) melting (Fig. 2E and 2F, Supplemental Fig. 4B-E, Supplemental Table 3). RPA^{E240K} and RPA^{V227A} showed a greater extent (Fig. 2E) and rate of melting of h-telG4 sequences (Fig. 2F) versus RPA^{WT}, especially at sub-saturating protein concentrations. We

conclude that all three mutants have a high affinity for DNA and RPA^{V227A} and RPA^{E240K} exert a higher affinity and telomere unfolding capacity than RPA^{WT}.

RPA1^{E240K} mutation results in premature telomere shortening and abnormal hematopoiesis in an iPSC model

To understand the effect of *RPA1* variants on telomere length regulation and hematopoiesis (Fig. 3A), we used CRISPR/Cas9-mediated mutagenesis to introduce homozygous E240K variant (denoted as *RPA1*^{E240K}) into the endogenous RPA1 locus of a healthy female donor iPSC line (*RPA1*^{WT}). *RPA1* whole gene sequencing and Sanger analysis (Fig. 3A) confirmed on-target mutagenesis. *RPA1*^{WT} and *RPA1*^{E240K} iPSC lines had similar *RPA1* protein expression (Fig. 3B), normal karyotype (Supplemental Fig. 5A) and intact pluripotency (Supplemental Fig. 5B). Given the role of RPA1 in replication and DNA repair, we assessed baseline cell cycle and activation of DNA damage response to irradiation, which showed similar cell cycle profiles and γ-H2AX signaling in both *RPA1*^{WT} and *RPA1*^{E240K} iPSCs (Supplemental Fig. 6A-B). We then measured telomere length using Q-FISH (Fig. 3C, Supplemental Fig. 6C) and TRF (Supplemental Fig. 6D), which revealed significantly shorter telomeres in *RPA1*^{E240K} compared to *RPA1*^{WT} iPSC independent of cell passage. Significant telomere length reduction was also observed in *RPA1*^{E240K} iPSC-derived HP (Fig. 3D). To assess whether RPA1^{E240K} affects hematopoiesis and whether *RPA1*^{E240K} iPSC model could faithfully recapitulate P1 phenotype of pancytopenia, we evaluated iPSC-derived hematopoietic differentiation with flow cytometric enumeration of cell subsets (Supplemental Fig. 7). Consistent with reduced hematopoietic potential, *RPA1*^{E240K} iPSC yielded lower percentage of cells with CD43^{+}CD45^{+} hematopoietic phenotype compared to *RPA1*^{WT} (Fig. 3E). Differentiation of HP cells (CD34^{+}CD43^{+}) to terminal lineages was also compromised. Specifically, *RPA1*^{E240K} produced significantly less CD71^{+}CD235^{+} erythroid (Fig. 3F) and CD45^{+}CD18^{+}CD11b^{+} myeloid populations (Fig. 3G) with cytological evidence of scarce orthochromatic proerythroblasts (Supplemental Fig. 8A-B) and large macrophage-like cells, respectively (Supplemental Fig. 8C-D).

Somatic inactivation of germline *RPA1* mutation results in benign clonal hematopoiesis with long-term potential

The atypical clinical course in P1 with stable blood counts over two decades (Fig. 4A) without progression of leukoplakia (Fig. 1C) and telomere shortening (Fig. 1D) led us to examine patient specimens for potential rescue mechanisms. We found that allelic frequency of *RPA1* c.718G>A, p.E240K variant is reduced in BM (27%) compared to fibroblast DNA (50%) (Fig. 4B). Two somatic events were identified in BM: a second-site truncating (stop-gain) *RPA1* mutation c.1735G>T, p.K579* at 10% allelic

frequency (Fig. 4B) and a uniparental isodisomy of chromosome 17p (UPD17p) (Fig. 4C). To understand the mechanism of these rescue events, we first confirmed the stop-gain c.1735G>T mutation to be *in cis* with germline c.718G>A allele (Fig. 4D, Supplemental Fig. 9). Bulk RNA sequencing of BM cells revealed total loss of the somatic mutation concurrent with reduction of germline variant to 13% (Fig. 4B). Ultradeep RNA sequencing performed to accurately quantify these mutations confirmed nearly absent expression (0.8%) of somatic c.1735G>T substitution (Fig. 4E). To assess whether the two mosaic events were sustained over time, we performed serial SNP-array and deep sequencing of hematopoietic cells. We also demonstrated a significant expansion of UPD17p clone over 15 years (Fig. 4C) which corresponds to the declining germline c.718G>A allele on serial BM evaluations (Fig. 4F). In addition, we observed an independent increase in clonal burden of somatic c.1735G>T, p.K579* mutation over time (Fig. 4F).

Analysis of clonality at single-cell resolution identifies mutually exclusive rescue events that arise in early hematopoiesis

To dissect the clonal architecture at single cell level, we interrogated P1 BM using scDNAseq. Three clones with unique *RPA1* allelic patterns were detected at age 13 years: heterozygous c.718G>A (native state hematopoiesis), homozygous wild type (rescue clone 1 with UPD17p), and heterozygous c.718G>A with concurrent somatic c.1735G>T (rescue clone 2 with stop-gain mutation) (Fig. 4G). This clonal pattern was confirmed in hematopoietic progenitor colonies derived from BM (Supplemental Fig. 10A). Clonal trajectories assessed over 4 years demonstrated minor expansion of rescue clone 2 at the expense of native hematopoiesis (Fig. 4G). Finally, we investigated clonal origins within hematopoietic hierarchy by using barcoded antibodies in the same reaction to track surface markers on single cell level. Both rescue clones were found within hematopoietic stem and progenitor cells (HSPC) characterized by the expression of CD34, CD90, and CD38 (Fig. 4H). In addition, compared to native hematopoiesis, these clones were enriched in $CD11b^+$ myeloid and $CD19^+$ B-cells but markedly reduced in $CD3^+$ T-lymphocytes (Fig. 4H). Similar skewing of UPD17p lesion towards myeloid lineage was observed comparing SNP arrays results in bulk granulocytes versus lymphocytes (Supplemental Fig. 10B). The enrichment of both somatic clones in $CD11^+$ and $CD19^+$ cells with a near absence of rescue hematopoiesis in T-lymphocytes (Fig. 4H) is likely due to their long lifespan compared to short-lived myeloid and B-cell lineages. Of note, no malignant somatic alterations (i.e. CHIP mutations) were found using WES in BM after diagnosis.

DISCUSSION

It was not until the late 20th century that *DKC1* was identified as the first gene to cause dyskeratosis congenita, a syndrome described in 1910 with features of nail dystrophy, oral leukoplakia and skin pigmentation anomalies[36]. Since then, an increasing number of genes involved in telomere homeostasis have been associated with classical phenotypes of TBD/STS which include BMF and mucocutaneous triad, but also manifestations such as isolated pulmonary fibrosis. Our study expands the genetic spectrum of this entity by describing three germline *RPA1* variants clustering to DBD-A to be associated with a novel Mendelian disorder that clinically resembles a TBD/STS with short telomeres. Consistent with the variable expressivity and penetrance observed in TBD/STS, the four affected individuals developed a wide range of disease features at various ages with the common denominator of telomere shortening. Constitutional manifestations included facial and eye anomalies present in two patients, pulmonary disease in two, and a classic mucocutaneous triad in one. Hematopoietic system was affected in three patients who had BMF, MDS, T- and B-cell lymphopenia and hypogammaglobulinemia. Only one patient (P3) did not have abnormal blood counts, thereby not prompting further marrow analysis. Similar to other TBD/STS manifestations reported in literature[37], it is possible that this patient's disease had slow progression in hematopoietic system while resulting in premature aging in the lungs manifesting as pulmonary fibrosis.

The well-established role of RPA1 in telomere maintenance using mammalian and non-mammalian model systems predated our discovery of germline *RPA1* mutations in patients with TBD/STS and short telomeres. Historically, many *S. cerevisiae rfa1 (RPA1* paralog) mutants demonstrated increased sensitivity from DNA damage, defective checkpoints, and gross chromosomal aberrations owing to RPA1 function in DNA replication and repair[18,19,38]. Smith et al. first showed the presence of RPA at the telomeric ends maximally during S phase, as well as, telomere shortening in a synergistic yku70-*rfa1-D228Y S. cerevisiae* model[38]. *rfa1-D228Y* was then modeled in *S. pombe* (paralog is *rad11-D223Y*), which confirmed sensitivity to UV and γ-irradiation and exhibited reduced telomere length, suggesting that RPA is directly involved in telomere maintenance[32]. Follow up studies further supported the role of RPA1 in telomere biology by demonstrating RPA1 ability to bind and unfold telomeric G-quadruplexes[26,28,39], regulate telomerase and telomerase access to chromosomal ends[24,40] and preventing accumulation of single strand telomeric DNA in alternative lengthening of telomeres (ALT) positive cells[29]. Furthermore, *rad11*-D223Y in *S. pombe*, corresponding to human *RPA1* D228Y, was shown to reduce binding affinity for telomeric ssDNA and G4 quadruplexes, resulting in telomere shortening[25-27]. Overall, multiple studies have established the role of RPA1 in telomere biology.

Unlike the previously studied variants, all patient RPA1 mutant proteins exhibit enhanced binding to ssDNA. Furthermore, *RPA1*^{V227A} and *RPA1*^{E240K} harbor increased binding affinity to telomeric sequences and a greater rate and extent of melting of G-quadruplexes whereas *RPA1*^{T270A} mutant was equivocal to *RPA1*^{WT}. Binding of RPA to DNA is dynamic and regulates loading of other essential proteins on ssDNA. It is plausible that GOF *RPA1*^{V227A} and *RPA1*^{E240K} mutations alter the access between chromosomal ends and telomere maintenance machinery such as shelterin or telomerase complex and/or challenge the RPA-to-POT1 switch and POT1 capping function, ultimately leaving telomeric ends vulnerable to telomere degrading transactions[41]. Additionally, specific point mutations might affect interactions of RPA1 with other proteins, which could explain why *RPA1*^{T270A} behaves different than other RPA1 mutants in binding to telomere sequences. The molecular mechanism by which RPA1 exerts telomere maintenance is highly complex and remains elusive, requiring further investigation.

To directly examine whether a single amino acid exchange in RPA1 can be deleterious to eukaryotic cells, we modeled *RPA1*^{E240K} variant in an iPSC model derived from a healthy donor. We deliberately chose to introduce patient *RPA1*^{E240K} mutation into a healthy donor iPSC cell line instead of generating patient derived iPSCs to eliminate the confounding effect of other unknown patient-specific mutations. The rationale for establishing a homozygous knock-in was to avoid the development of rescue events found in P1 and to exaggerate the biological phenotype for a gene variant that is associated with a late-onset disease. We were able to recapitulate the biological phenotype of significant telomere shortening in *RPA1*^{E240K} iPSC, as well as iPSC-derived hematopoiesis. Finally, we demonstrated that *RPA1*^{E240K} mutation resulted in reduced capacity to generate iPSC-derived HP and decreased erythroid/myeloid differentiation. This is in line with other reported iPSC short telomere disease models where hematopoietic insufficiency has been observed[42].

Stereotypic dyskeratosis congenita shows a progressive course with BMF leading to severe cytopenias over time[6]. Remarkably, over 18 years, P1 followed an atypical course for TBD/STS marked by stabilization of hematological features. Using this thread, we unraveled the unique propensity of germline RPA1^{E240K} mutation to facilitate the development of two independent somatic genetic escape lesions in P1. These were a second-site truncating *RPA1* mutation causing degradation of germline mutant RNA and a UPD17p recombination resulting in replacement of germline variant with a wild type allele. It seems that these mosaic events can expand and outcompete the native state *RPA1*-mutated hematopoiesis without signs of exhaustion or malignant transformation. Somatic genetic rescue in

hematopoiesis has been described in TBD/STS genes such as *TERC*, *TERT*, *TINF2*, and *DKC1*, which arise in response to LOF mutations[43-48]. The novelty of our finding lies in the identification of somatic rescue arising in response to a GOF RPA1^{E240K} mutation, a known phenomenon recently described in patients with GOF mutations in SAMD9/SAMD9L genes[49-51]. Additionally, we observed improvement of oral leukoplakia in P1 over time. One can speculate that mucosal tissue in this patient also underwent somatic reversion, given that somatic mosaicism has been recently shown to be common in healthy human tissues[52].

Our study does have potential limitations that should be noted. Firstly, due to the small size of our cohort, we are unable to characterize the full phenotypic spectrum associated with *RPA1* mutations. Secondly, telomere length was measured either by flow-FISH or TRF Southern blot, which was inherent to our patient cohort, presenting in four countries at different ages to unique providers with different strategies for sample banking. Although outside the scope of this study, further research efforts are required to understand the exact mechanism of how *RPA1* mutations cause telomere shortening, how RPA regulates telomere length, and how this compares to other telomere associated genes[53,54].

In summary, we identify *RPA1* mutations to be associated with telomere shortening in humans, which calls for careful consideration of *RPA1* missense variants in the workup of patients with TBD/STS phenotypes. Germline *RPA1* variants can either be permissive, as seen with RPA1^{V227A} that is associated with reduced penetrance or severely "hematotoxic" leading to somatic inactivation, as observed with RPA1^{E240K}. We speculate that germline *RPA1* alterations may be more common in human disease, given that somatic *RPA1* mutations occur in ~1% of cancers[55]. Additional efforts are needed to not only find further pathogenic *RPA1* variants but to also elucidate the role of *RPA1* in human telomere biology.

ACKNOWLEDGEMENTS

This work was supported by grants from the ERAPERMED GATA2-HuMo 2018-123, Deutsche Krebshilfe Max Eder Grant 109005, Fritz-Thyssen Foundation 10.17.1.026MN, and St Jude American Lebanese Syrian Associated Charities (ALSAC) (to MWW), José Carreras Leukämie-Stiftung (to VPL), BMBF MyPred 01GM1911A (to CMN, MWW, ME). Research reported in this manuscript was supported by the NIH Common Fund, through the Office of Strategic Coordination/Office of the NIH Director under Award Number U01HG007709. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The St. Jude Cancer Center Core Cytogenetics laboratory is supported by NIH, NCI (P30 CA21765-41) and ALSAC. MS is supported by R35GM131704 and P30CA086862SC. PR, CS, and CK are supported by the Agence Nationale de la Recherche (ANR-20-CE12-0012TeloRPA). HB and AK are supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany`s Excellence Strategy (EXC 22167-390884018). P3 exome study was supported by a grant from the Chancellerie des Universités de Paris (legs Poix). VG and PR laboratories are supported by the “Ligue Nationale Contre le Cancer” (LNCC) (Equipe labélisée). PR is a scientist from Centre National de la Recherche Scientifique (CNRS). SC is supported by Project Fondation ARC, Projet Emergence-Cancéropôle PACA. We thank the patients for participation and acknowledge our collaborators: Dirk Lebrecht, Marco Teller, Ali-Riza Kaya Wilfried Truckenmüller, Maria Siskou-Zwecker, Axel Gebert (Freiburg) for laboratory assistance and data management; Loizos Petrikkos and Kondylia Antoniadi (Athens) for patient management; Ibrahima Ba, Bruno Crestani (Marseille Cancer Research Centre) for technical assistance; Dr. Lindsay Burrage for exome analysis and Dr. Filiz Seeborg for P4 referral; Yawei Hui, Shibiao Wan, Yiping Fan, Gang Wu (St Jude Center for Applied Bioinformatics), OMICS computing cluster (University of Lübeck, Germany), Emmanuelle Olivier and Patrick Nitschké (Imagine Institute, Paris, France/ Université de Paris) for bioinformatics support; Robert Durruthy-Durruthy (MissionBio) for support on scDNAseq analysis; Amabel Orogo (Illumina) and Jordan Sheetz (Bio-Rad Laboratories) for technical support; Dr. Mihaela Onciu (St Jude Department of Pathology) for iPSC derived erythroid and myeloid cytology review; Dr. Aaron Taylor from St Jude Center of Imaging for imaging consultation; Drs. Sunita Dsouza, Maria Lillo Osuna and Min-Joon Han (St. Jude) for iPSC technical assistance; Virginia Valentine, and Julia Wilbourne (St Jude Cytogenetics core); and Drs. Mitchell Weiss and John Crispino for helpful discussions.

AUTHORSHIP CONTRIBUTION

RS, SSS, MWW, CS, SC, MSW, MS conceived and designed the experiments. RS, SSS, AK, HB, TCC, MB, VPL, JAR, CK, MWW performed genomic data analysis. RS, MH, SLG, CG, LS, FB, MBV, SMPM, AGF, DC, VG, CS, PR, MSW, MS, and SC performed and/or interpreted functional experiments. FB, SH, ML, CK, MC, SN, JAR, SP, CMN, ME, RS, MWW were involved in patient care, collecting clinical data and clinical testing. MWW, SC, MSW, MS supervised the experiments, and MWW oversaw study design. All authors contributed to the manuscript and approved of the final version.

COMPETING INTERESTS STATEMENT

The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing completed at Baylor Genetics Laboratories.

Appendix: study group members

The members of the Undiagnosed Diseases Network are the following:

Maria T. Acosta, Margaret Adam, David R. Adams, Pankaj B. Agrawal, Justin Alvey, Laura Amendola, Ashley Andrews, Euan A. Ashley, Mahshid S. Azamian, Carlos A. Bacino, Guney Bademci, Eva Baker, Ashok Balasubramanyam, Dustin Baldridge, Jim Bale, Michael Bamshad, Deborah Barbouth, Pinar Bayrak-Toydemir, Anita Beck, Alan H. Beggs, Edward Behrens, Gill Bejerano, Jimmy Bennet, Beverly Berg-Rood, Jonathan A. Bernstein, Gerard T. Berry, Anna Bican, Stephanie Bivona, Elizabeth Blue, John Bohnsack, Carsten Bonnenmann, Devon Bonner, Lorenzo Botto, Brenna Boyd, Lauren C. Briere, Elly Brokamp, Gabrielle Brown, Elizabeth A. Burke, Lindsay C. Burrage, Manish J. Butte, Peter Byers, William E. Byrd, John Carey, Olveen Carrasquillo, Ta Chen Peter Chang, Sirisak Chanprasert, Hsiao-Tuan Chao, Gary D. Clark, Terra R. Coakley, Laurel A. Cobban, Joy D. Cogan, Matthew Coggins, F. Sessions Cole, Heather A. Colley, Cynthia M. Cooper, Heidi Cope, William J. Craigen, Andrew B. Crouse, Michael Cunningham, Precilla D'Souza, Hongzheng Dai, Surendra Dasari, Joie Davis, Jyoti G. Dayal, Matthew Deardorff, Esteban C. Dell'Angelica, Katrina Dipple, Daniel Doherty, Naghmeh Dorrani, Argenia L. Doss, Emilie D. Douine, David D. Draper, Laura Duncan, Dawn Earl, David J. Eckstein, Lisa T. Emrick, Christine M. Eng, Cecilia Esteves, Marni Falk, Liliana Fernandez, Carlos Ferreira, Elizabeth L. Fieg, Laurie C. Findley, Paul G. Fisher, Brent L. Fogel, Irman Forghani, William A. Gahl, Ian Glass, Bernadette Gochuico, Rena A. Godfrey, Katie Golden-Grant, Madison P. Goldrich, David B. Goldstein, Alana Grajewski, Catherine A. Groden, Irma Gutierrez, Sihoun Hahn, Rizwan Hamid, Kelly Hassey, Nichole Hayes, Frances High, Anne Hing, Fuki M. Hisama, Ingrid A. Holm, Jason Hom, Martha Horike-Pyne, Alden Huang, Yong Huang, Laryssa Huryn, Rosario Isasi, Fariha Jamal, Gail P. Jarvik, Jeffrey Jarvik, Suman Jayadev, Lefkothea Karaviti, Jennifer Kennedy, Shamika Ketkar, Dana Kiley, Shilpa N. Kobren, Isaac S. Kohane, Jennefer N. Kohler, Deborah Krakow, Donna M. Krasnewich, Elijah Kravets, Susan Korrick, Mary Koziura, Joel B. Krier, Seema R. Lalani, Byron Lam, Christina Lam, Grace L. LaMoure, Brendan C. Lanpher, Ian R. Lanza, Lea Latham, Kimberly LeBlanc, Brendan H. Lee, Hane Lee, Roy Levitt, Richard A. Lewis, Sharyn A. Lincoln, Pengfei Liu, Xue Zhong Liu, Nicola Longo, Sandra K. Loo, Joseph Loscalzo, Richard L. Maas, John MacDowall, Ellen F. Macnamara, Calum A. MacRae, Valerie V. Maduro, Bryan C. Mak, May Christine V. Malicdan, Laura A. Mamounas, Teri A. Manolio, Rong Mao, Kenneth Maravilla, Thomas C. Markello, Ronit Marom, Gabor Marth, Beth A. Martin, Martin G. Martin, Julian A. Martínez-Agosto, Shruti Marwaha, Jacob McCauley, Allyn McConkie-Rosell, Alexa T. McCray, Elisabeth McGee, Heather Mefford, J. Lawrence Merritt, Matthew Might, Ghayda Mirzaa, Eva Morava, Paolo M. Moretti, Deborah Mosbrook-Davis, John J. Mulvihill, Mariko Nakano-Okuno, Avi Nath, Stan F. Nelson, John H. Newman, Sarah K. Nicholas, Deborah Nickerson, Shirley Nieves-Rodriguez, Donna Novacic, Devin Oglesbee, James P. Orengo, Laura Pace, Stephen Pak, J. Carl Pallais, Christina GS. Palmer, Jeanette C. Papp, Neil H. Parker, John A. Phillips III, Jennifer E. Posey, Lorraine Potocki, Bradley Power, Barbara N. Pusey, Aaron Quinlan, Wendy Raskind, Archana N. Raja, Deepak A. Rao, Genecee Renteria, Chloe M. Reuter, Lynette Rives, Amy K. Robertson, Lance H. Rodan, Jill A. Rosenfeld, Natalie Rosenwasser, Francis Rossignol, Maura Ruzhnikov, Ralph Sacco, Jacinda B. Sampson, Mario Saporta, C. Ron Scott, Judy Schaechter, Timothy Schedl, Kelly Schoch, Daryl A. Scott, Vandana Shashi, Jimann Shin, Rebecca Signer, Edwin K. Silverman, Janet S. Sinsheimer, Kathy Sisco, Edward C. Smith, Kevin S. Smith, Emily Solem, Lilianna Solnica-Krezel, Ben Solomon, Rebecca C. Spillmann, Joan M. Stoler, Jennifer A. Sullivan, Kathleen Sullivan, Angela Sun, Shirley Sutton, David A. Sweetser, Virginia Sybert, Holly K. Tabor, Amelia L. M. Tan, Queenie K.-G. Tan, Mustafa Tekin, Fred Telischi, Willa Thorson, Audrey Thurm, Cynthia J. Tifft, Camilo Toro, Alyssa A. Tran, Brianna M. Tucker, Tiina K. Urv, Adeline Vanderver, Matt Velinder, Dave Viskochil, Tiphanie P. Vogel, Colleen E. Wahl, Stephanie Wallace, Nicole M. Walley, Chris A. Walsh, Melissa Walker, Jennifer Wambach, Jijun Wan, Lee-kai Wang, Michael F. Wangler, Patricia A. Ward, Daniel

Wegner, Monika Weisz-Hubshman, Mark Wener, Tara Wenger, Katherine Wesseling Perry, Monte Westerfield, Matthew T. Wheeler, Jordan Whitlock, Lynne A. Wolfe, Jeremy D. Woods, Kim Worley, Shinya Yamamoto, John Yang, Muhammad Yousef, Diane B. Zastrow, Wadih Zein, Chunli Zhao, Stephan Zuchner, Hugo Bellen, and Rachel Mahoney.

REFERENCES

1. Blackburn EH, Epel ES, Lin J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science.* 2015;350(6265):1193-1198.
2. Niewisch MR, Savage SA. An update on the biology and management of dyskeratosis congenita and related telomere biology disorders. *Expert Rev Hematol.* 2019;12(12):1037-1052.
3. Schratz KE, Haley L, Danoff SK, et al. Cancer spectrum and outcomes in the Mendelian short telomere syndromes. *Blood.* 2020;135(22):1946-1956.
4. Calado RT, Young NS. Telomere diseases. *N Engl J Med.* 2009;361(24):2353-2365.
5. Dokal I, Vulliamy T, Mason P, Bessler M. Clinical utility gene card for: Dyskeratosis congenita - update 2015. *Eur J Hum Genet.* 2015;23(4).
6. Savage SA. Dyskeratosis Congenita. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews((R)).* Seattle (WA)1993.
7. Wold MS. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem.* 1997;66:61-92.
8. Bhat KP, Cortez D. RPA and RAD51: fork reversal, fork protection, and genome stability. *Nat Struct Mol Biol.* 2018;25(6):446-453.
9. de Laat WL, Appeldoorn E, Sugasawa K, Weterings E, Jaspers NG, Hoeijmakers JH. DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev.* 1998;12(16):2598-2609.
10. Krejci L, Altmannova V, Spirek M, Zhao X. Homologous recombination and its regulation. *Nucleic Acids Res.* 2012;40(13):5795-5818.
11. Li GM. Mechanisms and functions of DNA mismatch repair. *Cell Res.* 2008;18(1):85-98.
12. Caldwell CC, Spies M. Dynamic elements of replication protein A at the crossroads of DNA replication, recombination, and repair. *Crit Rev Biochem Mol Biol.* 2020;55(5):482-507.
13. Chen R, Wold MS. Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. *Bioessays.* 2014;36(12):1156-1161.
14. Dahai Y, Sanyuan S, Hong L, Di Z, Chong Z. A relationship between replication protein A and occurrence and prognosis of esophageal carcinoma. *Cell Biochem Biophys.* 2013;67(1):175-180.
15. Givalos N, Gakiopoulou H, Skliri M, et al. Replication protein A is an independent prognostic indicator with potential therapeutic implications in colon cancer. *Mod Pathol.* 2007;20(2):159-166.
16. Wang J, Yang T, Chen H, Li H, Zheng S. Oncogene RPA1 promotes proliferation of hepatocellular carcinoma via CDK4/Cyclin-D pathway. *Biochem Biophys Res Commun.* 2018;498(3):424-430.
17. Zhu Y, Yi Y, Bai B, et al. The silencing of replication protein A1 induced cell apoptosis via regulating Caspase 3. *Life Sci.* 2018;201:141-149.
18. Chen C, Kolodner RD. Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants. *Nat Genet.* 1999;23(1):81-85.
19. Chen C, Umezu K, Kolodner RD. Chromosomal rearrangements occur in S. cerevisiae rfa1 mutator mutants due to mutagenic lesions processed by double-strand-break repair. *Mol Cell.* 1998;2(1):9-22.
20. Haring SJ, Mason AC, Binz SK, Wold MS. Cellular functions of human RPA1. Multiple roles of domains in replication, repair, and checkpoints. *J Biol Chem.* 2008;283(27):19095-19111.
21. Wang Y, Putnam CD, Kane MF, et al. Mutation in Rpa1 results in defective DNA double-strand break repair, chromosomal instability and cancer in mice. *Nat Genet.* 2005;37(7):750-755.
22. Wang Y, Zhang W, Edelmann L, Kolodner RD, Kucherlapati R, Edelmann W. Cis lethal genetic interactions attenuate and alter p53 tumorigenesis. *Proc Natl Acad Sci U S A.* 2010;107(12):5511-5515.

23. Smith J, Rothstein R. A mutation in the gene encoding the Saccharomyces cerevisiae single-stranded DNA-binding protein Rfa1 stimulates a RAD52-independent pathway for direct-repeat recombination. *Mol Cell Biol.* 1995;15(3):1632-1641.
24. Luciano P, Coulon S, Faure V, et al. RPA facilitates telomerase activity at chromosome ends in budding and fission yeasts. *EMBO J.* 2012;31(8):2034-2046.
25. Audry J, Maestroni L, Delagoutte E, et al. RPA prevents G-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends. *EMBO J.* 2015;34(14):1942-1958.
26. Ray S, Qureshi MH, Malcolm DW, Budhathoki JB, Celik U, Balci H. RPA-mediated unfolding of systematically varying G-quadruplex structures. *Biophys J.* 2013;104(10):2235-2245.
27. Safa L, Delagoutte E, Petruseva I, et al. Binding polarity of RPA to telomeric sequences and influence of G-quadruplex stability. *Biochimie.* 2014;103:80-88.
28. Salas TR, Petruseva I, Lavrik O, et al. Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Res.* 2006;34(17):4857-4865.
29. Grudic A, Jul-Larsen A, Haring SJ, et al. Replication protein A prevents accumulation of single-stranded telomeric DNA in cells that use alternative lengthening of telomeres. *Nucleic Acids Res.* 2007;35(21):7267-7278.
30. Kibe T, Ono Y, Sato K, Ueno M. Fission yeast Taz1 and RPA are synergistically required to prevent rapid telomere loss. *Mol Biol Cell.* 2007;18(6):2378-2387.
31. Kobayashi Y, Sato K, Kibe T, et al. Expression of mutant RPA in human cancer cells causes telomere shortening. *Biosci Biotechnol Biochem.* 2010;74(2):382-385.
32. Ono Y, Tomita K, Matsuura A, et al. A novel allele of fission yeast rad11 that causes defects in DNA repair and telomere length regulation. *Nucleic Acids Res.* 2003;31(24):7141-7149.
33. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA.* 2014;312(18):1870-1879.
34. Sahoo SS, Pastor VB, Goodings C, et al. Clinical evolution, genetic landscape and trajectories of clonal hematopoiesis in SAMD9/SAMD9L syndromes. *Nat Med.* 2021.
35. Werner B, Beier F, Hummel S, et al. Reconstructing the in vivo dynamics of hematopoietic stem cells from telomere length distributions. *Elife.* 2015;4.
36. Heiss NS, Knight SW, Vulliamy TJ, et al. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet.* 1998;19(1):32-38.
37. Kam MLW, Nguyen TTT, Ngeow JYY. Telomere biology disorders. *NPJ Genom Med.* 2021;6(1):36.
38. Smith J, Zou H, Rothstein R. Characterization of genetic interactions with RFA1: the role of RPA in DNA replication and telomere maintenance. *Biochimie.* 2000;82(1):71-78.
39. Cohen S, Jacob E, Manor H. Effects of single-stranded DNA binding proteins on primer extension by telomerase. *Biochim Biophys Acta.* 2004;1679(2):129-140.
40. Schramke V, Luciano P, Brevet V, et al. RPA regulates telomerase action by providing Est1p access to chromosome ends. *Nat Genet.* 2004;36(1):46-54.
41. Flynn RL, Centore RC, O'Sullivan RJ, et al. TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature.* 2011;471(7339):532-536.
42. Winkler T, Hong SG, Decker JE, et al. Defective telomere elongation and hematopoiesis from telomerase-mutant aplastic anemia iPSCs. *J Clin Invest.* 2013;123(5):1952-1963.
43. Alder JK, Stanley SE, Wagner CL, Hamilton M, Hanumanthu VS, Armanios M. Exome sequencing identifies mutant TINF2 in a family with pulmonary fibrosis. *Chest.* 2015;147(5):1361-1368.
44. Gutierrez-Rodrigues F, Donaires FS, Pinto A, et al. Pathogenic TERT promoter variants in telomere diseases. *Genet Med.* 2019;21(7):1594-1602.
45. Jongmans MC, Verwiel ET, Heijdra Y, et al. Revertant somatic mosaicism by mitotic recombination in dyskeratosis congenita. *Am J Hum Genet.* 2012;90(3):426-433.
46. Maryoung L, Yue Y, Young A, et al. Somatic mutations in telomerase promoter counterbalance germline loss-of-function mutations. *J Clin Invest.* 2017;127(3):982-986.

47. Perdigones N, Perin JC, Schiano I, et al. Clonal hematopoiesis in patients with dyskeratosis congenita. *Am J Hematol.* 2016;91(12):1227-1233.
48. Revy P, Kannengiesser C, Fischer A. Somatic genetic rescue in Mendelian haematopoietic diseases. *Nat Rev Genet.* 2019;20(10):582-598.
49. Buonocore F, Kuhnen P, Suntharalingham JP, et al. Somatic mutations and progressive monosomy modify SAMD9-related phenotypes in humans. *J Clin Invest.* 2017;127(5):1700-1713.
50. Chen DH, Below JE, Shimamura A, et al. Ataxia-Pancytopenia Syndrome Is Caused by Missense Mutations in SAMD9L. *Am J Hum Genet.* 2016;98(6):1146-1158.
51. Narumi S, Amano N, Ishii T, et al. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. *Nat Genet.* 2016;48(7):792-797.
52. Yizhak K, Aguet F, Kim J, et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science.* 2019;364(6444).
53. Fok WC, Shukla S, Vessoni AT, et al. Posttranscriptional modulation of TERC by PAPD5 inhibition rescues hematopoietic development in dyskeratosis congenita. *Blood.* 2019;133(12):1308-1312.
54. Shukla S, Jeong HC, Sturgeon CM, Parker R, Batista LFZ. Chemical inhibition of PAPD5/7 rescues telomerase function and hematopoiesis in dyskeratosis congenita. *Blood Adv.* 2020;4(12):2717-2722.
55. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 2019;47(D1):D941-D947.

TABLES

Patient features				Germline mutation				Disease features		
Patient	**Gender**	**Age of onset (years)**	***RPA1* status**	**Exon**	**VAF% (TD)**	**Population databases**	**CADD**	**Hematopoietic**	**Other**	**Recent status**
P1	F	10	c.718G>A, p.E240K *de novo*	9	24%[#] (58)	absent	23.3	Refractory cytopenia of childhood	Mucocutaneous triad Congenital eye anomaly*	No interventions Stable blood counts at age 28 years
P2	F	13	c.680T>C, p.V227A	8	51% (118)	gnomADv3.1: 76078 genomes (1/152156 alleles) TOPMed freeze 8: 132345 genomes (2/264690 alleles)	26.9	Myelodysplastic syndrome with excess blasts	Mild restrictive lung disease Facial dysmorphism**	HSCT Death at age 14 years (GVHD, infection, pulmonary fibrosis)
P3	F	58			53% (49)			BM not examined CBC normal	Familial pulmonary fibrosis***	Progression of pulmonary fibrosis on antifibrotics
P4	F	birth	c.808A>G, p.T270A *de novo*	10	41% (51)	absent	20.9	BM not examined Low TRECs Lymphopenia Hypogammaglobulinemia	Prematurity (33 weeks)	IgG replacement Stable CBC at age 3 years

Table 1: Genomic and clinical features of patients with germline heterozygous *RPA1* variants. F, female; TD, total depth; gnomAD v3.1 and TOPMed Bravo freeze 8 population databases with population genome sequencing data; CADD, combined annotation dependent depletion, pathogenicity threshold of 15; DANN, deleterious annotation of genetic variants using neural networks, range of 0 to 1 with higher values with greater probability of being damaging; BM, bone marrow; CBC, complete blood count; TRECs, T-cell receptor excision circles; HSCT, hematopoietic stem cell transplantation; GVHD, graft versus host disease.

Confirmed germline in fibroblasts, decreased VAF is due to UPD in blood cells; *Congenital persistent hyperplastic primary vitreous of the left eye requiring enucleation; **Bilateral blepharophimosis, epicanthus inversus, eyelid ptosis, and thick eyebrows with high arch found in P2 and carrier father; *** Pulmonary fibrosis diagnosed in two sisters of P3 (Supplemental Fig. 2). NM_002945.5 was used for variant nomenclature.

FIGURE LEGENDS

Figure 1. Clinical characteristics of patients with identified *RPA1* variants. (A) Germline *RPA1* variants identified in four pedigrees using exome sequencing. Black filled, open dotted and open symbols denote affected individuals with heterozygous (het) *RPA1* mutations, unaffected carriers, and unaffected family members without mutations, respectively. (B) Top panel: Schematic of the *RPA1* gene with 17 coding exons illustrated as black lines and untranslated regions shown in red (NM_002945.5). The three unique patient variants are located on exons 8, 9 and 10. Bottom panel: Human RPA1 protein illustration with four oligonucleotide/oligosaccharide-binding (OB) fold domains F, A, B, and C with amino acid boundaries shown below. OB folds A, B and C are single strand (ss) DNA binding domains. Red arrows indicate location of 3 *RPA1* missense alterations. (C) Bone marrow (BM) findings in patient cohort (top panel): P1 BM aspirate smears (Wright-Giemsa staining) at time of clinical presentation at age 10 years showing marked reduction in cell content with hypoplasia of all lineages (left image) and megaloblastic maturation of erythroid precursors (middle image). P2 BM infiltrated with myeloblasts (right image) consistent with myelodysplastic syndrome with excess blasts. Pulmonary findings by chest computed tomography (CT) in patient cohort (middle panel): P2 chest CT (left image) during hospitalization demonstrating necrotizing pneumonitis with diffuse ground-glass opacities and large air-filled cavities with differential diagnosis including pulmonary GVHD, opportunistic infection(s), pulmonary fibrosis, or a combination of the above conditions. P3 presented at age 58 years with cough and exertional dyspnea with chest CT (middle image) findings showing intralobular reticulations with traction bronchiectasis and mild honeycombing, with left asymmetric and basal, subpleural predominance. At 61 years, P3 chest CT (right image) showed significant progression of asymmetric lung fibrosis with left predominance and massive basal honeycombing. Mucocutaneous abnormalities in P1 (bottom panel): oral leukoplakia at 19 years of age (top left image) with mild improvement at 25 years of age (top right image), nail dystrophy (bottom left image) and reticular skin pigmentation on the ventral neck (bottom right image). (D) Telomere length analysis by flow cytometry-based-fluorescence in-situ hybridization (flow-FISH) was carried out in lymphocytes of P1 (red circles) and P4 (blue circles). P1 telomere length is less than 1st percentile at 24 and 27 years of age (red circles). P4 telomere length is at the 5th percentile at 1.75 years of age and less than 1st percentile at age 3.33 years (blue circles). All measurements were carried out in triplicates, and mean telomere length was calculated in kilobases (kb) in relation to the internal control (bovine thymocytes) with known telomere length. 356 healthy controls used for calculation of the 1st (solid line, bottom), 5th (dashed line, bottom), 95th (dashed line, top), and 99th (solid line, top) percentile curves. (E) Telomere restriction fragment (TRF) analysis in peripheral blood DNA from P2 and family and P3 compared to healthy age matched control (Ctrl), digested with HinfI and RsaI enzymes followed by separation on 0.7% agarose gel.

Figure 2. All RPA mutant heterotrimeric proteins exhibit increased affinity for ssDNA and RPA^{V227A} and RPA^{E240K} possess increased capacity to unfold telomeric G-quadruplex (h-telG4) DNA. (A) Schematic depiction of the experimental schemes for the FRET-based assays. Binding of RPA:RPA1^{WT} (RPA^{WT}), RPA:RPA1^{V227A} (RPA^{V227A}), RPA:RPA1^{V270A} (RPA^{V270A}) and RPA:RPA1^{E240K} (RPA^{E240K}) proteins was monitored using 1 nM dT30 ssDNA molecules (top), 1 nM dT15 or TTAGGGTAAGGGTAA telomeric DNA sequence (middle) labeled with Cy3 and Cy5 fluorescent dyes at the 5' and 3' ends, respectively. High FRET corresponds to free ssDNA, while low FRET reflects RPA binding. Unfolding of the telomeric G-

quadruplex (h-telG4) was monitored using $(TTAGGG)_5$ sequence. FRET between the Cy3 and Cy5 dyes calculated for the h-telG4 in the absence of proteins and in buffer containing K^+ corresponds to 100% folded quadruplex, while the FRET value of h-telG4 in the presence of saturating concentrations of RPA^{E240K} in buffer containing Li^+ corresponds to 100% unfolded h-telG4. (B) Stoichiometric binding (1 RPA: 1 dT30 molecule) was observed for RPA^{E240K} (orange) and nearly stoichiometric binding was observed for the RPA^{WT} (black), RPA^{V227A} (pink) and RPA^{V270A} (blue). The arrows mark the respective protein concentrations at inflection points of the two-line linear regression fit. (C) dT15 and (D) TTAGGGTAAGGGTAA telomeric DNA sequence binding to RPA^{WT} (black), RPA^{V227A} (pink), RPA^{V270A} (blue) and RPA^{E240K} (orange). The data were fitted to quadratic binding equation. The calculated K_ds with respective fitting errors are listed in Supplemental Table 3. (E) and **(**F), Melting of the h-telG4 DNA, stabilized by the presence of 100 mM KCl. (E) Extents of the h-telG4 melting reactions were calculated from the plateaus of each respective time course. (F) h-telG4 melting rates for RPA^{WT} (black), RPA^{V227A} (pink), RPA^{V270A} (blue) and RPA^{E240K} (orange) were calculated from the slopes of FRET change during the first 20 seconds of each time course (Supplemental Fig. 4). The data were fitted to quadratic binding equation. The calculated apparent K_ds with respective fitting errors are listed in Supplemental Table 3. In all panels the data are shown as average for three independent experiments. Error bars represent standard deviation. Where not shown, error bars are smaller than the data points.

Figure 3. Human *RPA1*^{E240K} iPSC demonstrate telomere shortening and reduced hematopoietic potential. (A) Left panel: Healthy control iPSC ($RPA1^{WT}$) following CRISPR/Cas9-guided homozygous c.718G>A (p.E240K, as E240K) modification within endogenous *RPA1* locus (*RPA1*^{E240K}) confirmed with Sanger analysis. Right panel: Illustration of iPSC monolayer-based differentiation to hematopoietic progenitors and subsequently to erythroid and myeloid cell lineages. Day 10 HP cells were FACS-sorted and cultured in erythroid or myeloid differentiation media for 14 days. In parallel, iPSC-derived HP were further cultured until day 21 to assess for expression of pan-hematopoietic markers. (B) Immunoblot analysis of RPA1 expression in *RPA1*^{WT} and *RPA1*^{E240K} iPSC whole cell extracts with histone-H3 as loading control. (C) Telomere length in *RPA1*^{WT} passage 17 and $RPA1^{E240K}$ passage 12 iPSCs (D) and iPSC-derived HP cells using quantitative fluorescence in situ hybridization (Q-FISH). Graphs represent mean ± SEM of one of three independent experiments (****P < 0.0001; Student t-test). (E) Decreased percentage of $CD43^+CD45^+$ *RPA1*^{E240K} hematopoietic cells compared to *RPA1*^{WT} at days 16 and 21. Data represent mean ± SEM of two independent experiments (*P = 0.03, **P = 0.0074; Student t-test). (F) Graphical representation of $CD71^+CD235^+$ erythroid cells from iPSC-derived erythroid cultures at day 14. Data represents mean ± SEM of 4 independent experiments (***P = 0.0002; Student t-test). **I**, Plot representation of $CD45^+CD11b^+$ myeloid cells from iPSC cultures. Data represent mean ± SEM of 2 independent experiments (**P = 0.0018; Student t-test).

Figure 4. Natural evolution of disease and somatic genetic rescue in patient 1. (A) White blood cells (WBC, triangles), hemoglobin (Hgb, circles) and platelets (diamonds) plotted over 23 years for patient 1 (P1, pedigree 1). Red arrow indicates time of clinical presentation. (B) Illustration of germline *RPA1* variant in exon 9 and somatic mutation in exon 16 with respective DNA Sanger electropherograms from bone marrow. Bottom table depicts variant allelic frequencies from exome sequencing performed in bone marrow and skin fibroblast DNA and RNA sequencing from bone marrow at 20 years of age. (C)

Copy number neutral uniparental isodisomy (UPD) encompassing *RPA1* locus at 17p13.3 (red arrow) identified using single nucleotide polymorphism (SNP) array. Serial SNP-array analysis in bone marrow granulocytes demonstrates UPD expansion over time (denoted by purple brackets). (D) Schematic of *RPA1* locus (grey bar) with germline (c.718) and somatic (c.1735) mutational spots 17kb apart. Three haplotype orientations between c.718 and c.1735 identified in marrow DNA of P1 at age 19 years from 2 independent experiments using digital droplet PCR: left haplotype, wt/wt (c.718G wild type/c.1735G wild type) denoted by black boxes; middle haplotype, mut/wt (c.718A mutant/c.1735G wild type) denoted by green and black boxes; right haplotype, mut/mut (c.718A mutant/c.1735T mutant) denoted by green and red boxes. (E) Ultradeep amplicon sequencing of bone marrow DNA and RNA targeting position of *RPA1* somatic mutation (c.1735) confirms near total loss of mutant RNA. (F) Longitudinal deep sequencing in bone marrow samples from diagnosis to age 25 years showing decrease in allele frequency of the germline c.G718G>A variant (red line) and increase of the somatic c.1735G>T mutation (blue line). (G) Single cells from P1 bone marrow at ages 13 and 17 were sequenced for germline (*RPA1*:chr17:1782314:G>A) and somatic (*RPA1*:chr17:1798378:G>T) mutational positions using single cell (sc) DNA sequencing Tapestri platform. Violin plot shows 3 clonal populations including, homozygous wild type (blue, $RPA1^{WT/WT}$, rescue clone 1 = UPD17p), heterozygous $RPA1^{E240K/WT}$ (gold, native state hematopoiesis) and heterozygous c.718G>A with concurrent c.1735G>T stop-gain (red, $RPA1^{E240K/WT + K579*}$= rescue clone 2). (H) Tapestri single cell multi-omic analysis combining DNA mutation data and surface protein expression performed in P1 bone marrow at age 17 years. Panels depict 3 clones (color coding identical to panel G) constructed from 2,110 high quality cells with normalized protein expression of markers for hematopoietic stem and progenitor cells (CD34), stem cells (CD90), progenitors (CD38), and terminally differentiated cells including, myeloid (CD11b), B-lymphoid (CD19) and T-lymphoid (CD3) cells.