

Somatic reversion impacts MDS/AML evolution in the short telomere syndromes

Kristen E. Schratz^{1,5,6}, Valeriya Gaysinskaya¹, Zoe L. Cosner¹, Emily A. DeBoy^{1,2},
Zhimin Xiang¹, Laura Kasch-Semenza², Liliana Florea^{2,3}, Pali D. Shah³, Mary Armanios^{1,2,4,5,6}

Departments of Oncology¹, Genetic Medicine², Medicine³ and Molecular Biology and Genetics⁴
Telomere Center at Johns Hopkins⁵
Sidney Kimmel Comprehensive Cancer Center⁶
Johns Hopkins University School of Medicine
Baltimore, MD 21287

Correspondence
Mary Armanios, M.D.
1650 Orleans St.
CRB 1 Room 186
Baltimore, MD 21287
410-502-3817 phone
410-502-6240 fax
marmani1@jhmi.edu

Abstract

Inherited mutations in telomerase and other telomere maintenance genes manifest in the premature aging short telomere syndromes. Myelodysplastic syndromes and acute myeloid leukemia (MDS/AML) account for 75% of associated malignancies, but how these cancers overcome the germline telomere defect is unknown. We used targeted ultra-deep sequencing to detect candidate somatic reversion mutations hypothesizing they may promote MDS/AML evolution. While no controls carried somatic mutations in telomere maintenance genes (0 of 28), 29% of adults with germline telomere maintenance defects carried at least one (16 of 56, $P < 0.001$). In addition to *TERT* promoter mutations which were present in 19%, we identified *POT1* and *TERF2IP* mutations in 13%. *POT1* mutations impaired telomere binding and in some cases were identical to those found in familial melanoma associated with longer telomere length. Exclusively in patients with germline defects in telomerase RNA (TR), we identified somatic mutations in nuclear RNA exosome genes, *RBM7*, *SKIV2L2*, and *DIS3*, where loss-of-function upregulates mature TR levels. Paradoxically, somatic reversion events were more prevalent in patients who were MDS/AML-free ($P = 0.01$, RR 5.0, 95% CI 1.4-18.9), and no MDS/AML patient had more than one mutant clone ($P = 0.048$). Our data identify diverse somatic adaptive mechanisms in the short telomere syndromes, and raise the possibility that their presence alleviates the telomere crisis that promotes transformation to MDS/AML.

Germline mutations in telomerase and other telomere-related genes are thought to be one of if not the most common known genetic cause of adult-onset myelodysplastic syndromes and possibly acute myeloid leukemia (MDS/AML)(1-4). In line with this observation, in patients with inherited mutations in telomerase and other telomere maintenance genes, MDS/AML are the most common cancers(5, 6). They account for 75% of short telomere syndrome malignancies but the lifetime risk is 10%(5, 6). The vast majority of short telomere syndrome MDS/AML cases are age-related; they manifest after the age of 50 often concurrently with idiopathic pulmonary fibrosis or other telomere-mediated pulmonary disease(5-8). In primary cells, critically short telomeres provoke a DNA damage response that triggers a p53-dependent checkpoint that signals apoptosis or cellular senescence(9, 10). This checkpoint mediates the degenerative phenotypes of the short telomere syndromes which manifest most prominently as bone marrow failure and pulmonary fibrosis(11, 12). How these myeloid cancers overcome the background short telomere checkpoint to sustain their replicative potential remains unknown. Moreover, the factors that predict which patients go on to develop MDS/AML are not understood. The latter question is particularly timely since pulmonary fibrosis patients with short telomere syndromes are now identifiable prior to lung transplantation, but have a significantly increased risk of acquiring post-transplant MDS and AML(6).

Here, we used an ultra-deep sequencing approach to test the hypothesis that acquired clonal somatic mutations that offset the inherited short telomere defect arise in telomere-mediated MDS/AML. As we show, in contrast to our expectations, we found diverse somatic reversion mechanisms that show great specificity relative to the germline defect. Surprisingly, however, these somatic mutations were overall rare in MDS/AML patients and had a higher prevalence

and allele burdens in patients who were MDS/AML-free. These data have implications for fundamental understanding of how somatic adaptation in the hematopoietic system may provide protection against MDS/AML evolution in at least some inherited bone marrow failure syndromes.

Results

To understand the mechanisms of telomere maintenance in short telomere MDS/AML patients, we compared the prevalence of candidate somatic mutations among 84 individuals who were over the age 40 and who were divided across three groups: healthy controls (mean age 64, n=28,), short telomere syndrome patients who did not have MDS/AML (mean age 60, n=40) and short telomere syndrome patients who fulfilled World Health Organization criteria for MDS/AML (mean age 56, n=16). Figure 1A summarizes their clinical characteristics. Healthy controls had normal telomere length, while the other two groups had similarly abnormal short length defects (Figure 1B, C). We used a customized ultra-deep error-corrected sequencing platform(13, 14), and designed our experiment to detect low frequency somatic mutations down to 0.5% variant allele frequency (VAF). Using this pipeline, we targeted variants in 16 candidate genes in which mutations could functionally promote telomere maintenance. Their functions were grouped in four categories: 1) promoters of the core telomerase components, telomerase reverse transcriptase and telomerase RNA, *TERT* and *TR*, 2) the six shelterin genes of which four are mutated in cancer-prone families hypothesized to have long telomere syndromes(15, 16), 3) six nuclear RNA exosome genes involved in *TR* processing and where loss-of-function has been shown to increase mature TR levels, and 4) *ATRX* and *DAXX*, known to be mutated in cancers that rely on alternative lengthening of telomeres (ALT) mechanisms (Supplementary Table 1).

In addition, we examined *TP53* since its loss may allow bypass of the short telomere checkpoint(9). Among the three groups, we identified 36 clonal somatic mutations: 34 single nucleotide variants (SNVs) and 2 insertions/deletions. They fell in six of the 17 candidate genes (35%): *TP53*, the *TERT* promoter, *POT1*, *TERF2IP* and *RBM7*, *SKIV2L2* and *DIS3* (Figure 2A-D and Table 1). Only one was found in a healthy control; and that mutation was a low allele frequency *TP53* (VAF 1%) subclinical clonal hematopoiesis mutation. Another 10 *TP53* missense mutations fell predominantly in the DNA binding domain, and all are recurrently mutated in cancer (Figure 2A). The latter were divided equally among short telomere patients with and without MDS/AML (Figure 2A).

All the remaining somatic telomere gene mutations fell exclusively in short telomere syndrome patients with 30% carrying at least one (16 of 56 short telomere vs. 0 of 28 controls, $P < 0.001$, Fisher's exact test) (Table 1). The most common mutations were in the *TERT* promoter; they were seen at the -146 and -124 hotspots upstream of *TERT*'s transcription start site (Figure 2B). Nineteen percent (11 of 56) of the combined short telomere patients carried at least one of these mutation; this rate is higher than the 5-7% prevalence previously reported in patients with pulmonary fibrosis and dyskeratosis congenita which utilized less sensitive sequencing methods(17, 18). These *TERT* promoter mutations are identical to somatically mutated nucleotides found in numerous cancers and which create *de novo* ETS-binding sites that upregulate transcription of the intact *TERT* allele in *trans* with the germline mutation(17, 19-21).

The next most common class of mutations fell in *POT1*, a negative regulator of telomere elongation. Two mutations fell in canonical splice junction sequences (Figure 2C and Table 1).

For *POT1*'s and the missense mutations in the other four protein-coding genes, all but one had Combined Annotation-Dependent Depletion (CADD) scores ≥ 20 signifying they represent the top 1st percentile of damaging mutations, and five had a score of ≥ 30 indicating they were in the top 0.1 percentile of damaging mutations (Table 1). Several of the shelterin gene mutations, and mutations in nuclear RNA exosome genes, have been documented to be also somatically mutated in cutaneous malignant melanoma (CMM), chronic lymphocytic leukemia (CLL) among other cancers (Table 1). Somatic mutations at the POT1 p.H266 residue, for example, are the most common telomere maintenance mechanism mutated in CLL(22, 23), and two mutations, POT1 p.T41A and c.1687-1G>T have been identified in familial melanoma, a long telomere syndrome phenotype(15, 24) and DeBoy and Armanios, unpublished). A third group of mutations fell in RNA processing genes, and the impact of loss-of-function of the nuclear RNA exosome targeting (NEXT) complex genes, *RBM7*, *SKIV2L2*, and the nuclear RNA exosome essential nuclease, *DIS3*, on increasing TR levels has been previously documented(25, 26). Given the high frequency of *POT1* missense mutations, we tested their functional consequences in an electrophoretic mobility shift assay (EMSA). The four missense mutations in the OB folds fell within 3.5 Angstrom's of telomeric DNA in the POT1-telomere crystal structure(23, 27) and showed impaired DNA binding to varying degrees (Figure 2E-G). Notably, the K33E mutation phenocopied a mutant that was deleted for the entire first OB fold of POT1 (Figure 2E-G). In contrast, POT1 p.C605S, which fell in the TPP1-interacting domain (Figure 1C), showed no DNA binding defects. These collective data supported that multiple *de novo* somatic mutations arise relatively commonly in telomerase and telomere gene mutation carriers. They are functionally impactful and overlap with somatic telomere maintenance mutations seen in cancer.

We next examined if the telomere-related somatic mutations clustered in the MDS/AML patients, but found a surprising and paradoxical result. Short telomere adults who were MDS/AML-free had more mutations than the MDS/AML patients (RR 5.0, 95% CI 1.4-18.9, $P=0.01$, Fisher's exact test). Among adults without MDS/AML, 35% had ≥ 1 and 18% had ≥ 2 somatic mutations (14 and 7 of 40, respectively, Figure 3A). But for MDS/AML patients, only 13% (2 of 16) carried ≥ 1 and none carried ≥ 2 mutations ($P=0.1$ and $P=0.08$, respectively). The somatic mutant clones were also on average 3.8-fold larger in MDS/AML-free adults (mean 14.7% vs. 3.9%). Notably, none of the MDS/AML patients had a clonal somatic mutation in a telomere gene $>10\%$ VAF ($P=0.048$, Fisher's exact test) (Figure 3B), suggesting this 10% threshold may be a useful predictive threshold if validated in other studies. We assessed the incidence of MDS/AML in the patients who carried somatic reversion mutations and all of them remained MDS/AML-free with a median follow-up of 2.0 years (range 3 months-11 years).

If these somatic mutations correlate negatively with MDS/AML onset, we would predict they would be limited to or enriched in the myeloid lineage. To test this directly, we used droplet digital PCR probes to quantify the mutant fraction in whole blood and compared it to myeloid and T cell fractions. We studied four mutations (*POT1* and *DIS3*) from three patients and found they were all enriched in the myeloid lineage while being low or undetectable in the T cell fraction (Figure 4). To definitively test if the somatic mutations we identified were indeed reversion events, we correlated the germline and somatic mutant pathways and found remarkable associations. While *TERT* promoter mutations predominated in germline *TERT* mutation carriers (8 of 11), *POT1* and *TERF2IP* mutations were associated with a more heterogeneous spectrum of germline mutations including *TERT*, *TINF2*, *RTEL1* and *PARN* (Table 1). This observation is

consistent with POT1 loss-of-function mutations playing a role in improving telomerase access or processivity(15). By contrast, mutations in the nuclear RNA exosome genes were exclusively detected in patients with germline defects in TR abundance or function. All three patients who carried these somatic mutations carried germline mutations in *TR* itself, the ribonucleoprotein *DKC1*, or had abnormally low *TR* levels (Table 1). These data support a model where myeloid-specific adaptation that arises in the highly replicative environment of hematopoiesis, and under the selective pressures of a germline short telomere defect, is associated with an MDS/AML-free state.

Discussion

In the high turnover environment of hematopoiesis, the selective pressures imposed by inherited mutations allow advantageous clonal mutations to persist(28). Nearly all the known somatic reversion mechanisms are direct in that they repair the germline mutant gene itself such as by correcting a frameshift or through homologous recombination with the *wild-type* allele(28-30). Here, we identified several indirect somatic reversion mechanisms in the short telomere syndromes, and show they overlap with known telomere maintenance mechanisms in cancer. Their high specificity, such as is the case for the nuclear exosome components in *TR* mutation carriers, underscores a remarkable versatility of the stressed hematopoietic system to acquire *de novo* adaptations. We did not observe gross measurable telomere lengthening effects (Supplementary Figure 1A), but recurrent mutations, often in the same genes or pathways within a given individual patient, highlight the competitive advantage they provide under the selective pressures of the short telomere environment. The finding of *POT1* recurrent mutations across multiple germline defects in *TERT* and other mutation carriers support a model where *POT1*

loss-of-function facilitates telomere elongation by improving telomerase access and/or processivity. Of note, several of the mutations we identified have been seen in cancer prone families with melanoma. The shared spectrum of mutations we describe here with somatic mutations seen in cancer as well as in germline mutant cancer prone families support a model wherein *POT1* mutations themselves do not promote genome instability but instead favor a state of telomere maintenance(15).

To our surprise, the adaptive mechanisms we identified, including mutations in the *TERT* promoter, had a higher prevalence in patients who did not have MDS/AML. These data shed light into how the short telomere bone marrow failure state may evolve to myeloid malignancy. One possibility is that short telomere progenitors undergoing crisis undergo fate-determining events that lead to a functional reversion or a maladaptive mechanism such as loss of chromosome 7. In the majority of cases in our study, these events were mutually exclusive. Comparing the 33% prevalence of reversion mutations we identified in adult short telomere patients (Figure 4), relative to the 10% overall incidence of MDS/AML, somatic reversion appears to be a more common and preferred adaptive mechanism. It is also possible that somatic reversion mutations may be similarly associated with an MDS/AML-free state in other MDS/AML inherited bone marrow failure states.

Our data raise the possibility that deep sequencing for at least some of the common mutations we identified may be a useful biomarker for assessing MDS/AML risk in some settings. IPF is the most common indication of lung transplantation world-wide and it is estimated that 35% of familial pulmonary fibrosis patients carry a germline defect in a telomere maintenance gene(31,

32). In the setting of lung transplantation, short telomere patients are exposed to cytotoxic medications that add selective pressures on hematopoiesis, and the presence of sizable clonal reversion mutations may provide a protective advantage. The median follow-up for patients with telomere-related somatic reversion mutations was two years. Additional studies and longer follow-up will be needed to test the predictive utility of somatic reversion testing in clinical settings.

Materials and Methods

Subjects and study approval. Subjects were recruited 2003 to 2020 as part of the Johns Hopkins Telomere Syndrome Study(6, 33). They were enrolled if they carried a pathogenic mutation in a telomere-related gene *and* had a personal/family history of a short telomere phenotype, had classic familial short telomere syndrome as defined(7), *or* had sporadic disease with low telomerase RNA levels(34). The MDS/AML diagnosis was assigned using WHO classification(35). Healthy controls were recruited as previously and were selected for having normal telomere length(36). The research was approved by Johns Hopkins Medicine's Institutional Review Board, and subjects gave written informed consent.

Germline sequencing, functional studies and telomere length measurement. Germline sequencing was performed on peripheral blood DNA using whole genome, exome, or targeted sequencing(25, 30, 36). Germline mutations not reported in Alder *et al.*(36) or Schratz *et al.*(6) are listed in Supplementary Table 1. Telomerase RNA quantification and *DKC1* functional analyses have been detailed(25, 37). Telomere length was measured by flow cytometry and fluorescence in situ hybridization (flowFISH) as described(36, 38).

Capture Design, Library preparation and Sequencing. We designed 150 bp read length probes (SureDesign) targeting 37 kb across 17 genes including two promoters (Supplementary Table 2). Libraries were prepared following the manufacturer's instructions (Agilent Haloplex^{HS} for 1-500kb target region). Briefly, 400 ng of genomic DNA diluted to 14.4 ng/uL were digested in eight restriction enzyme reactions and hybridized to the custom biotinylated HaloPlex^{HS} probe library incorporating sample indexes and a degenerate 10-nucleotide unique molecular barcode (i.e. UMI, to identify each individual captured DNA molecule) to facilitate base calling accuracy and quantification. Targets were ligated to form circularized fragments, captured using streptavidin beads and polymerase chain reaction amplified (24 cycles). Target libraries were validated using Fragment Analyzer followed by quantification of fragment smear analysis of 190bp to 640bp range. All samples were subsequently pooled at 4 nM concentration. A final AMPure XP magnetic bead cleanup was performed on the pooled library to eliminate any adapter-dimer traces.

Haloplex library sequencing. Sequencing was performed on an Illumina HiSeq 2500 using a 150 bp paired-end protocol, and data were demultiplexed following standard protocols. Reads were trimmed and aligned to the reference hg19 using BWA MEM in the Agilent Alissa Align and Call v.1.1.2.2 platform. Of 86 samples designed for Haloplex platform sequencing, 80 eventually fulfilled quality control criteria (3 failed at library preparation, 3 had low coverage). The mean depth of coverage was $12,522x \pm 4,591$ SD with a median of $10,952x \pm 5,959$ SD. The coverage depth for the *TERT* promoter was relatively lower and was calculated separately using Samtools (version 1.10, using htslib 1.10). Within the targeted interval (chr5:1295085-1295385, hg19), its coverage was limited to (ch5:1295085-1295115, hg19) and (chr5:1295224-1295385) and these intervals had a mean coverage of $2,457x \pm 1,204$ SD with a median 2,421x.

Three additional MDS/AML patients were consecutively recruited after the Haloplex sequencing protocol started and their DNA was sequenced using a clinically validated dual-indexed targeted method that applied a background error rate analysis to allow robust detection of somatic mutations with allele frequency $\geq 1\%$ (6, 39). This pipeline included the 3 most commonly mutated genes in the Haloplex analysis: *TP53*, *TERT promoter*, *POT1*, as well as *TINF2*. Supplementary Figure 1B lists the mutation types at the base pair level and shows 58% of SNVs were cytosine to thymidine transitions.

Variant calling and analysis. Variant calling was performed using Agilent Alissa Align and Call v1.1.2.2 software and annotated using ANNOVAR v4162018. To utilize the Haloplex error-corrected method of somatic variation, we first collapsed reads originating from the same sample molecule into read families and duplicates were removed to create consensus reads. Variants were filtered to include only those present in ≥ 3 unique read families. To exclude sequencing artifacts, we used a quality score of >90 and filtered for a variant score threshold of >0.3 in Alissa. Variants that were present in $>5\%$ of controls were additionally excluded as alignment artifacts. Variants with allele frequencies of 40-60% or $>95\%$ were deemed germline and excluded. We finally filtered for variants with minor allele frequencies >0.0005 in gnomAD v2.1.1 (any VAF), and these were excluded as they were thought to be possibly germline (performed November 1, 2020). The *TERT* promoter was manually analyzed and curated separately to include variants present in all reads of one or more consensus families at the canonical c.-124 and c.-146 positions irrespective of variant allele frequency. All of the variants reported were manually verified in Integrative Genomics Viewer (IGV).

Variant interpretation. Protein schema were drawn based on the longest isoform curated in the Gene Database (National Center for Biotechnology Information, accessed 10/21/20). Linear protein structures were visualized in SnapGene (v5.1.6), and conserved domains constructed were based on Uniprot (release 2020_05) and the literature. Variants were analyzed for their presence in the COSMIC database (v.92) and Combined Annotation Dependent Depletion (CADD) score (v1.6)(40).

POT1 gel shift binding assay. To test the functional significance of POT1 missense mutations, we cloned Myc-FLAG tagged POT1 into T7 expression vector (pcDNA 5/FRT, Addgene) and performed site-directed mutagenesis. A control with POT1 ^{Δ OB} was generated by deleting the first OB fold (amino acids 127-635). T7 expression vectors were used in an in vitro translation reaction using TNT Quick Coupled Transcription/Translation System (Promega), according to manufacturer's instructions. Briefly, a 50 μ l reaction containing 1 μ g of expression vector, 20 μ M methionine, and 40 μ l TNT Quick Master mix containing rabbit reticulocyte lysate was incubated at 30°C for 90 minutes and used immediately for the gel shift assay.

The POT1 binding assay were preformed using Odyssey Infrared EMSA Kit (LI-COR). Briefly, a 20 μ l reaction mixture was prepared containing 2 μ l of the IVT product, 10 nM IRDye800-labeled single stranded telomeric oligonucleotide, 2.5 μ M nonspecific ssDNA and 1 μ g poly(dI-dC), in binding buffer containing 100 mM Tris, 500 mM KCl, 12.5 mM DTT, 10 mM EDTA, and 0.25% Tween 20. We used this telomere oligonucleotide, 5'GGTTAGGGTTAGGGTTAGGG, and the non-specific ssDNA oligonucleotide 5'TTAATTAACCCGGGGATCCGGCTTGATCAACGAATGATCC as per Bauman *et al.*(41).

Reactions were incubated for 30 minutes at room temperature and protein-DNA complexes were resolved by gel electrophoresis on a 10% polyacrylamide Tris-Borate EDTA at 100V for 2h at 4°C. Gels were visualized using Li-Cor Odyssey Imaging System. Quantification was done using ImageJ (2.1.0).

Western blot. Using the NuPAGE SDS-PAGE gel system, 1 µl of in vitro translation reaction was run on 8% Bis-Tris gels with MOPS-SDS running buffer at 150 V and transferred to a PVDF membrane using the iBlot 2 Dry Blotting System on setting P0 (Thermo Fisher Scientific). Membranes were blocked for 1h before antibody staining. Membranes were stained first with anti-POT1 antibody (rabbit, NB500-176, 1:500, Novus) and subsequently with anti-rabbit horseradish peroxidase secondary antibody (HRP, Goat, 1:20,000, Cell Signaling Technology) before visualization via chemiluminescence (SuperSignal West Pico Plus Chemiluminescent substrate kit) using ImageQuant. Membranes were then stripped for 15 minutes, using New Blot IR Stripping buffer, and blocked again for 1h before staining with anti-Myc antibody (mouse, clone 4A6, 1:1000; Millipore). Membranes were subsequently stained with anti-mouse IRDye secondary antibody as appropriate (IR680, donkey, 1:10,000, LI-COR) before visualization on an Odyssey scanner (LI-COR). Quantification was done using ImageJ (2.1.0).

Cell fractionation. Aliquots of fresh blood or frozen ficoll-separated peripheral blood cells were fractionated using EasySep Human Myeloid Positive Selection and EasySep Human T Cell Enrichment kits (StemCell Technologies). DNA was extracted from the fractionated subsets using MagMAX DNA Multi-Sample Ultra 2.0 Kit (ThermoFisher) and quantified using Qubit dsDNA HS Assay kit (ThermoFisher).

Droplet Digital PCR. Droplet digital PCR (ddPCR) was performed according to manufacturer's specifications on a QX200 system (Bio-Rad). Sequence-specific primer and probes were designed using the Bio-Rad digital assay design engine: *POT1* p.K33E (dHsaMDS726207399), *POT1* p.M159V (dHsaMDS784505022), *POT1* p.H266Y (dHsaMDS401027863), *DIS3* p.D488N (HsaMDS2512362). An annealing temperature of 58°C was used except for the *POT1* p.K33E (54°C) and *DIS3* p.D488N (54°C). The wildtype and mutant probes were conjugated with HEX and FAM reporters, respectively. PCR amplicons were sequence-verified by cloning the products (TOPO TA Cloning Kit, Invitrogen). For each target, the mean number of droplets analyzed was $11,890 \pm 2,081$ SD and all reactions had >9,000 events analyzed. Data were analyzed in QX Manager (version 1.2-STD, Bio-Rad).

Statistics. Graphs were generated and statistical calculations were performed using GraphPad Prism 8.0.

Author contributions. KES designed the study and analyzed the clinical and genetic data. KES and VG selected the candidate genes. VG prepared the libraries with LKS. ZLC performed the POT1 functional studies and analyzed ddPCR data. EAD did computational analyses. ZX performed the fractionation and ddPCR experiments. LF supported designing the sequencing analysis pipeline. PDS analyzed clinical data. MA oversaw the project and wrote the paper with KES. All the authors reviewed and approved the manuscript.

Acknowledgements. We would like to acknowledge helpful conversations Dustin Gable, and technical support from the Johns Hopkins Genomics team, the Johns Hopkins Genetics Resources Core Facility and the Bloomberg School of Public Health Flow Cytometry Core. We are grateful for technical support from the Agilent team of Josh Zhiyong Wang, Gregory Miles and Weiwei Liu. This work was supported by RO1CA225027 and RO1HL119476, the Commonwealth Foundation, S&R Foundation Kuno Award, and the Williams Foundation (MA), T32HL007525, a grant from the Vera and Joseph Dresner Foundation and MacMillan Pathway to Independence Award (KES), F32HL142207 (VG), the Research Program for Medical Students (ZLC), T32GM007309 and the Turock Scholars Fund to the Telomere Center at Johns Hopkins (KES, EAD). The authors would also like to acknowledge a gift in the name of Mrs. P. Godrej.

Conflict of Interest. The authors declare no conflicts of interest.

References

1. Keel SB, Scott A, Sanchez-Bonilla M, Ho PA, Gulsuner S, Pritchard CC, et al. Genetic features of myelodysplastic syndrome and aplastic anemia in pediatric and young adult patients. *Haematologica*. 2016;101(11):1343-50.
2. Calado RT, Regal JA, Hills M, Yewdell WT, Dalmazzo LF, Zago MA, et al. Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. *Proc Natl Acad Sci U S A*. 2009;106(4):1187-92.
3. Reilly CR, Myllymaki M, Redd R, Padmananban S, Karunakaran D, Tesmer V, et al. The clinical and functional effects of TERT variants in myelodysplastic syndrome. *bioRxiv*. 2021.
4. Bluteau O, Sebert M, Leblanc T, Peffault de Latour R, Quentin S, Lainey E, et al. A landscape of germ line mutations in a cohort of inherited bone marrow failure patients. *Blood*. 2018;131(7):717-32.
5. Schratz KE, and Armanios M. Cancer and myeloid clonal evolution in the short telomere syndromes. *Curr Opin Genet Dev*. 2020;60:112-8.
6. Schratz KE, Haley L, Danoff SK, Blackford A, DeZern A, Gocke CD, et al. Cancer spectrum and outcomes in the Mendelian short telomere syndromes. *Blood*. 2020.
7. Parry EM, Alder JK, Qi X, Chen JJ, and Armanios M. Syndrome complex of bone marrow failure and pulmonary fibrosis predicts germline defects in telomerase. *Blood*. 2011;117(21):5607-11.
8. Merck SJ, and Armanios M. Shall we call them "telomere-mediated"? Renaming the idiopathic after the cause is found. *Eur Respir J*. 2016;48(6):1556-8.
9. Chin L, Artandi SE, Shen Q, Tam A, Lee SL, Gottlieb GJ, et al. p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. *Cell*. 1999;97(4):527-38.
10. Hemann MT, Rudolph KL, Strong MA, DePinho RA, Chin L, and Greider CW. Telomere dysfunction triggers developmentally regulated germ cell apoptosis. *Mol Biol Cell*. 2001;12(7):2023-30.
11. Alder JK, Barkauskas CE, Limjunyawong N, Stanley SE, Kembou F, Tudor RM, et al. Telomere dysfunction causes alveolar stem cell failure. *Proc Natl Acad Sci U S A*. 2015;112(16):5099-104.
12. Armanios M, and Blackburn EH. The telomere syndromes. *Nature reviews Genetics*. 2012;13(10):693-704.
13. Xia J, Miller CA, Baty J, Ramesh A, Jotte MRM, Fulton RS, et al. Somatic mutations and clonal hematopoiesis in congenital neutropenia. *Blood*. 2018;131(4):408-16.
14. Duncavage EJ, Jacoby MA, Chang GS, Miller CA, Edwin N, Shao J, et al. Mutation Clearance after Transplantation for Myelodysplastic Syndrome. *N Engl J Med*. 2018;379(11):1028-41.
15. McNally EJ, Luncsford PJ, and Armanios M. Long telomeres and cancer risk: the price of cellular immortality. *J Clin Invest*. 2019;130.
16. Stanley SE, and Armanios M. The short and long telomere syndromes: paired paradigms for molecular medicine. *Curr Opin Genet Dev*. 2015;33:1-9.
17. Maryoung L, Yue Y, Young A, Newton CA, Barba C, van Oers NS, et al. Somatic mutations in telomerase promoter counterbalance germline loss-of-function mutations. *J Clin Invest*. 2017;127(3):982-6.

18. Gutierrez-Rodrigues F, Donaires FS, Pinto A, Vicente A, Dillon LW, Cle DV, et al. Pathogenic TERT promoter variants in telomere diseases. *Genet Med*. 2019;21(7):1594-602.
19. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, and Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. *Science*. 2013;339(6122):957-9.
20. Bell RJ, Rube HT, Kreig A, Mancini A, Fouse SF, Nagarajan RP, et al. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science*. 2015.
21. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, et al. TERT promoter mutations in familial and sporadic melanoma. *Science*. 2013;339(6122):959-61.
22. Herling CD, Klaumunzer M, Rocha CK, Altmuller J, Thiele H, Bahlo J, et al. Complex karyotypes and KRAS and POT1 mutations impact outcome in CLL after chlorambucil-based chemotherapy or chemoimmunotherapy. *Blood*. 2016;128(3):395-404.
23. Ramsay AJ, Quesada V, Foronda M, Conde L, Martinez-Trillos A, Villamor N, et al. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat Genet*. 2013;45(5):526-30.
24. Shen E, Xiu J, Lopez GY, Bentley R, Jalali A, Heimberger AB, et al. POT1 mutation spectrum in tumour types commonly diagnosed among POT1-associated hereditary cancer syndrome families. *J Med Genet*. 2020;57(10):664-70.
25. Gable DL, Gaysinskaya V, Atik CC, Talbot CC, Jr., Kang B, Stanley SE, et al. ZCCHC8, the nuclear exosome targeting component, is mutated in familial pulmonary fibrosis and is required for telomerase RNA maturation. *Genes Dev*. 2019;33(19-20):1381-96.
26. Tseng CK, Wang HF, Burns AM, Schroeder MR, Gaspari M, and Baumann P. Human Telomerase RNA Processing and Quality Control. *Cell Rep*. 2015;13(10):2232-43.
27. Lei M, Podell ER, and Cech TR. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nature structural & molecular biology*. 2004;11(12):1223-9.
28. Revy P, Kannengiesser C, and Fischer A. Somatic genetic rescue in Mendelian haematopoietic diseases. *Nature reviews Genetics*. 2019;20(10):582-98.
29. Jongmans MC, Verwiel ET, Heijdra Y, Vulliamy T, Kamping EJ, Hehir-Kwa JY, et al. Revertant somatic mosaicism by mitotic recombination in dyskeratosis congenita. *Am J Hum Genet*. 2012;90(3):426-33.
30. Alder JK, Stanley SE, Wagner CL, Hamilton M, Hanumanthu VS, and Armanios M. Exome sequencing identifies mutant TINF2 in a family with pulmonary fibrosis. *Chest*. 2015;147(5):1361-8.
31. Silhan LL, Shah PD, Chambers DC, Snyder LD, Riise GC, Wagner CL, et al. Lung transplantation in telomerase mutation carriers with pulmonary fibrosis. *Eur Respir J*. 2014;44(1):178-87.
32. Stanley SE, Chen JJ, Podlevsky JD, Alder JK, Hansel NN, Mathias RA, et al. Telomerase mutations in smokers with severe emphysema. *J Clin Invest*. 2015;125(2):563-70.
33. Jonassaint NL, Guo N, Califano JA, Montgomery EA, and Armanios M. The gastrointestinal manifestations of telomere-mediated disease. *Aging cell*. 2013;12(2):319-23.
34. Stanley SE, Gable DL, Wagner CL, Carlile TM, Hanumanthu VS, Podlevsky JD, et al. Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. *Science translational medicine*. 2016;8(351):351ra107.

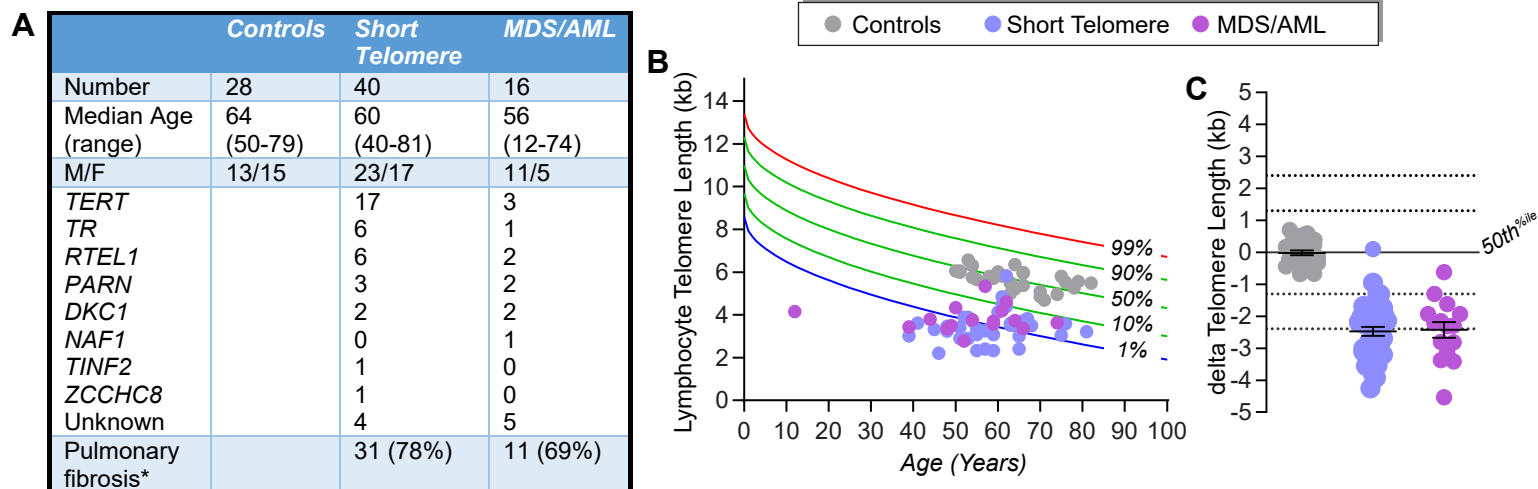
35. Swerdlow SH, and International Agency for Research on Cancer. WHO. *WHO classification of tumours of haematopoietic and lymphoid tissues*. Lyon: International Agency for Research on Cancer; 2017.
36. Alder JK, Hanumanthu VS, Strong MA, DeZern AE, Stanley SE, Takemoto CM, et al. Diagnostic utility of telomere length testing in a hospital-based setting. *Proc Natl Acad Sci U S A*. 2018;115(10):E2358-e65.
37. Gaysinskaya V, Stanley SE, Adam S, and Armanios M. Synonymous Mutation in DKC1 Causes Telomerase RNA Insufficiency Manifesting as Familial Pulmonary Fibrosis. *Chest*. 2020.
38. Baerlocher GM, Vulto I, de Jong G, and Lansdorp PM. Flow cytometry and FISH to measure the average length of telomeres (flow FISH). *Nat Protoc*. 2006;1(5):2365-76.
39. Zheng G, Chen P, Pallavajjala A, Haley L, Gondek L, Dezern A, et al. The diagnostic utility of targeted gene panel sequencing in discriminating etiologies of cytopenia. *Am J Hematol*. 2019;94(10):1141-8.
40. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, and Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-5.
41. Baumann P, Podell E, and Cech TR. Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing. *Mol Cell Biol*. 2002;22(22):8079-87.

Table 1. Germline and somatic mutations in telomere-related genes, their VAF and CADD score

	Germline Mutant Gene	Somatic Mutant Gene	Somatic Coding Variant	Somatic Protein Change	VAF %	CADD Score	COSMIC or Familial cancer
TERT promoter	<i>TERT</i>	<i>TERT</i>	c.-124 C>T		1.4%		CMM, Numerous cancers
	<i>TERT</i>	<i>TERT</i>	c.-124 C>T		1.6%		
	<i>TERT</i>	<i>TERT</i>	c.-124 C>T		2.8%		
	<i>NAF1</i>	<i>TERT</i>	c.-124 C>T		6.3%		
	<i>TERT</i>	<i>TERT</i>	c.-124 C>T		6.5%		
	<i>TERT</i>	<i>TERT</i>	c.-146 C>T		7.7%		
	<i>Unknown</i>	<i>TERT</i>	c.-124 C>A		23.8%		
	<i>TR</i>	<i>TERT</i>	c.-124 C>T		25.9%		
	<i>TERT</i>	<i>TERT</i>	c.-124 C>T		46.2%		
	<i>TERT</i>	<i>TERT</i>	c.-124 C>T		55.9%		
	<i>TERT</i>	<i>TERT</i>	c.-146 C>T		59.5%		
Shelterin	<i>Low TR</i>	<i>POT1</i>	c.T1813A	p.C605S	0.95%	26	-
	<i>TERT</i>	<i>POT1</i>	c.A97G	p.K33E	0.97%	26	CLL
	<i>Unknown</i>	<i>POT1</i>	c.1687-1G>T	<i>Splicing</i>	1.4%	33	Familial CMM
	<i>TERT</i>	<i>TERF2IP</i>	c.A1012C	p.K338Q	1.4%	26	-
	<i>RTEL1</i>	<i>POT1</i>	c.125-1G>A	<i>Splicing</i>	5.0%	34	ATCL
	<i>TINF2</i>	<i>POT1</i>	c.C796T	p.H266Y	6.1%	27	CMM, CLL
	<i>TERT</i>	<i>POT1</i>	c.A121G	p.T41A	13.3%	23	Familial CMM
	<i>TINF2</i>	<i>POT1</i>	c.A475G	p.M159V	14.6%	16	-
Nuclear RNA exosome	<i>TR</i>	<i>SKIV2L2</i>	c.G2561A	p.R854H	0.67%	32	-
	<i>TR</i>	<i>SKIV2L2</i>	c.G1052A	p.R351Q	0.73%	23	Gastric
	<i>Low TR</i>	<i>RBM7</i>	c.461delA	p.M155Cfs*49	2.4%	32	-
	<i>DKC1</i>	<i>DIS3</i>	c.G1600A	p.E534K	22.6%	27	-
	<i>TR</i>	<i>DIS3</i>	c.G1462A	p.D488N	24.4%	31	CLL, Numerous

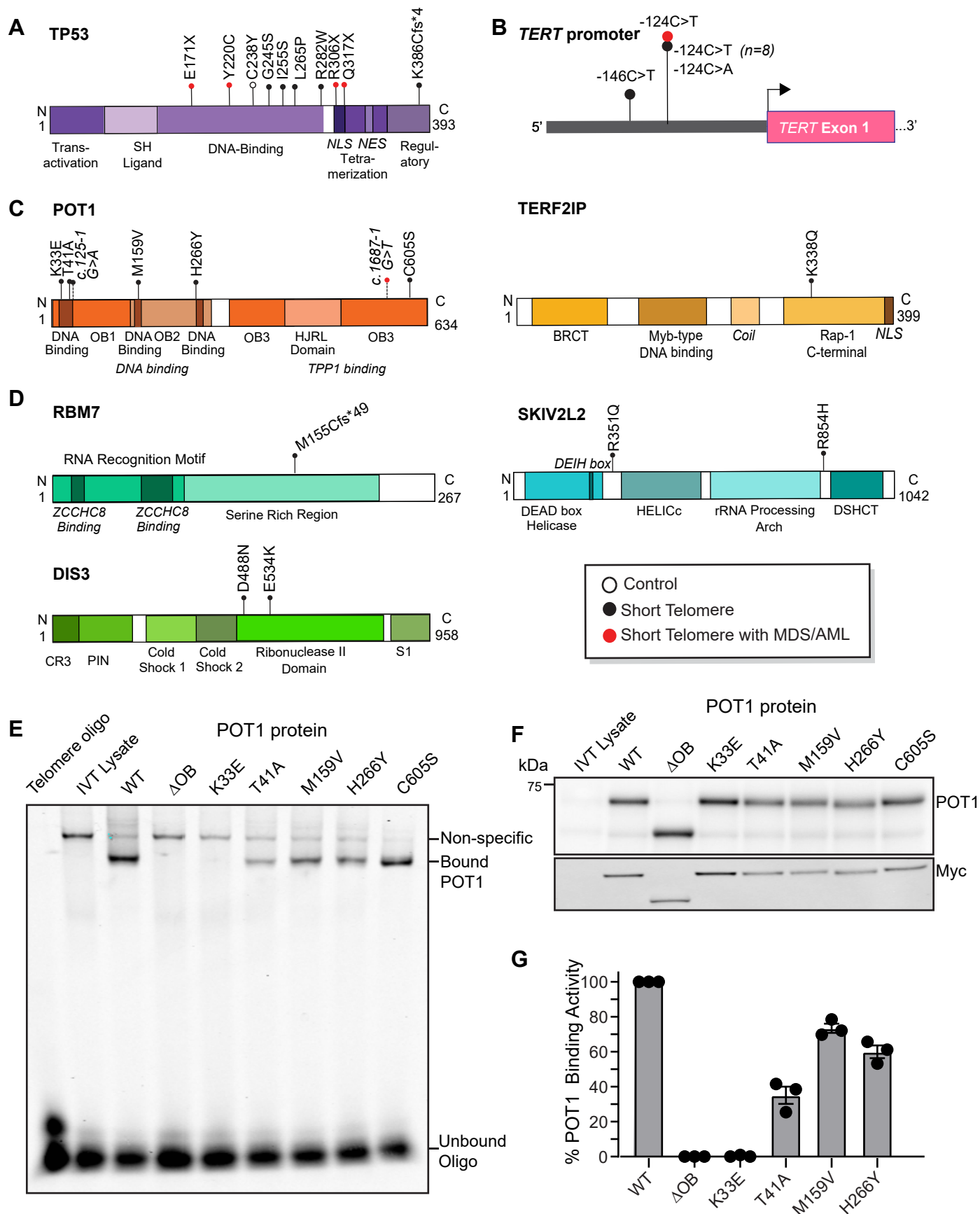
Abbreviations/Notes: CADD score refers to score from the Combined Annotation-Dependent Depletion (v1.6) tool. A higher CADD score denotes a more deleterious prediction with ≥ 20 referring to the top 1% of most deleterious possible substitutions, and ≥ 30 referring to the top 0.1%. ATCL refers to Adult T cell lymphoma; CMM refers to cutaneous malignant melanoma; CLL refers to chronic lymphocytic leukemia familial melanoma. VAF refers to variant allele frequency.

Figure 1



Clinical characteristics of controls, short telomere syndrome patients with and without myelodysplastic syndrome and acute myeloid leukemia (MDS/AML). (A) Demographics and mutant germline genes of controls and short telomere syndrome subjects. All the autosomal genes listed were heterozygous mutant. The unknown category included one patient with low telomerase RNA levels each. Pulmonary fibrosis refers to patients who had this concurrent diagnosis at the time of study recruitment and blood draw. (B) Telogram showing lymphocyte telomere length by flow cytometry and fluorescence in situ hybridization (flowFISH) relative to age-adjusted nomogram, and (C) is a plot of the difference of an individual lymphocyte telomere length relative to the median for age. Error bars represent standard error of the mean.

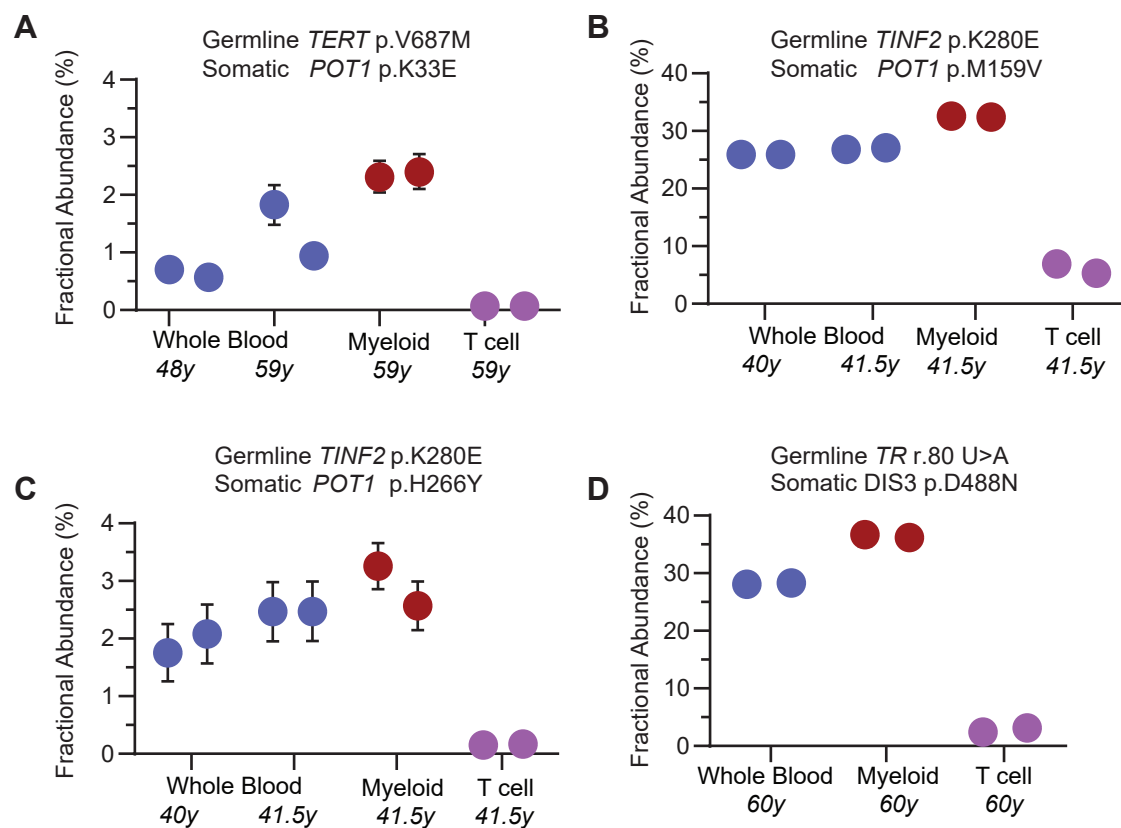
Figure 2



Somatic mutations relative to protein functional domains and POT1 functional studies.

(A), (B), (C) and **(D)** Schema showing somatic mutations relative to mutant protein/gene with conserved domains drawn for TP53, the *TERT* promoter, shelterin components and nuclear RNA exosome components, respectively. The key denotes the group in which the mutation was identified with open circle referring to controls, black circles to short telomere patients without MDS/AML, and red circles to short telomere patients with MDS/AML. **(E)** Electrophoretic mobility shift assay (EMSA) for POT1 mutants examining the binding capacity of each of the missense mutations identified. POT1^{ΔOB} refers to a control deleted for the first DNA binding domain (aa127-635). WT refers to *wild-type*. **(F)** Immunoblot of in vitro translated products shows missense mutant POT1 is stable with the Myc blot confirming specificity. **(G)** Mean intensity of binding relative to WT with error bars representing standard error of the mean. The data shown are from three EMSA reactions derived from two independent in vitro translations.

Figure 4



Fractional abundance of mutant clone size in leukocyte fractions quantified by droplet digital PCR (ddPCR). (A-D) ddPCR abundance of mutations are shown at study enrollment, and for three of the mutations at follow-up. The germline and somatic mutations are shown above, and the age at the time of draw is annotated below. DNA for ddPCR analysis was derived from whole blood, the myeloid fraction (CD33⁺ CD66⁺) and CD3⁺ T cell fraction. Each dot refers to mean fractional abundance from the Poisson distribution with error bars representing the minimum and maximum of the dataset with each reaction performed twice denoted. Where not seen, min/max lines fall within the boundaries of the graphed dot.