Poly(A)-specific ribonuclease (PARN) mediates 3′-end maturation of the telomerase RNA component

Diane H Moon[1–4], Matthew Segal[1–4], Baris Boyraz[1–5], Eva Guinan[6], Inga Hofmann[1,3,4], Patrick Cahan[1], Albert K Tai[7] & Suneet Agarwal[1–4]

Mutations in the *PARN* gene (encoding poly(A)-specific ribonuclease) cause telomere diseases including familial idiopathic pulmonary fibrosis (IPF) and dyskeratosis congenita[1,2], but how PARN deficiency impairs telomere maintenance is unclear. Here, using somatic cells and induced pluripotent stem cells (iPSCs) from patients with dyskeratosis congenita with *PARN* mutations, we show that PARN is required for the 3′-end maturation of the telomerase RNA component (TERC). Patient-derived cells as well as immortalized cells in which PARN is disrupted show decreased levels of *TERC*. Deep sequencing of *TERC* RNA 3′ termini shows that PARN is required for removal of post-transcriptionally acquired oligo(A) tails that target nuclear RNAs for degradation. Diminished *TERC* levels and the increased proportion of oligo(A) forms of *TERC* are normalized by restoring PARN, which is limiting for *TERC* maturation in cells. Our results demonstrate a new role for PARN in the biogenesis of TERC and provide a mechanism linking *PARN* mutations to telomere diseases.

PARN is a widely expressed cap-dependent poly(A) deadenylase with a canonical role in regulating global mRNA levels during development[3–6] and additional, more specialized functions, including end-trimming of the Dicer-independent microRNA miR-451 (ref. 7) and deadenylation of small nucleolar RNAs (snoRNAs)[8]. *PARN* loss-of-function mutations have recently been implicated in IPF[1] and dyskeratosis congenita[2], but why a deficiency in PARN, which has no known role in telomere biology, should mimic telomere diseases is unexplained[9].

TERC serves as the RNA template and scaffold for the telomerase reverse-transcriptase holoenzyme[10–12], and its levels are tightly controlled, thereby limiting telomerase activity and telomere elongation in the cell[13,14]. The *TERC* gene is amplified in several malignancies[15,16], whereas reduction in TERC levels due to genetic mutations in *DKC1*, *NOP10*, *NHP2* or *TERC* itself results in telomere disease[17–21]. Because of its 3′ box H/ACA domain[22], we hypothesized that TERC is regulated by PARN in a manner similar to that for other box H/ACA
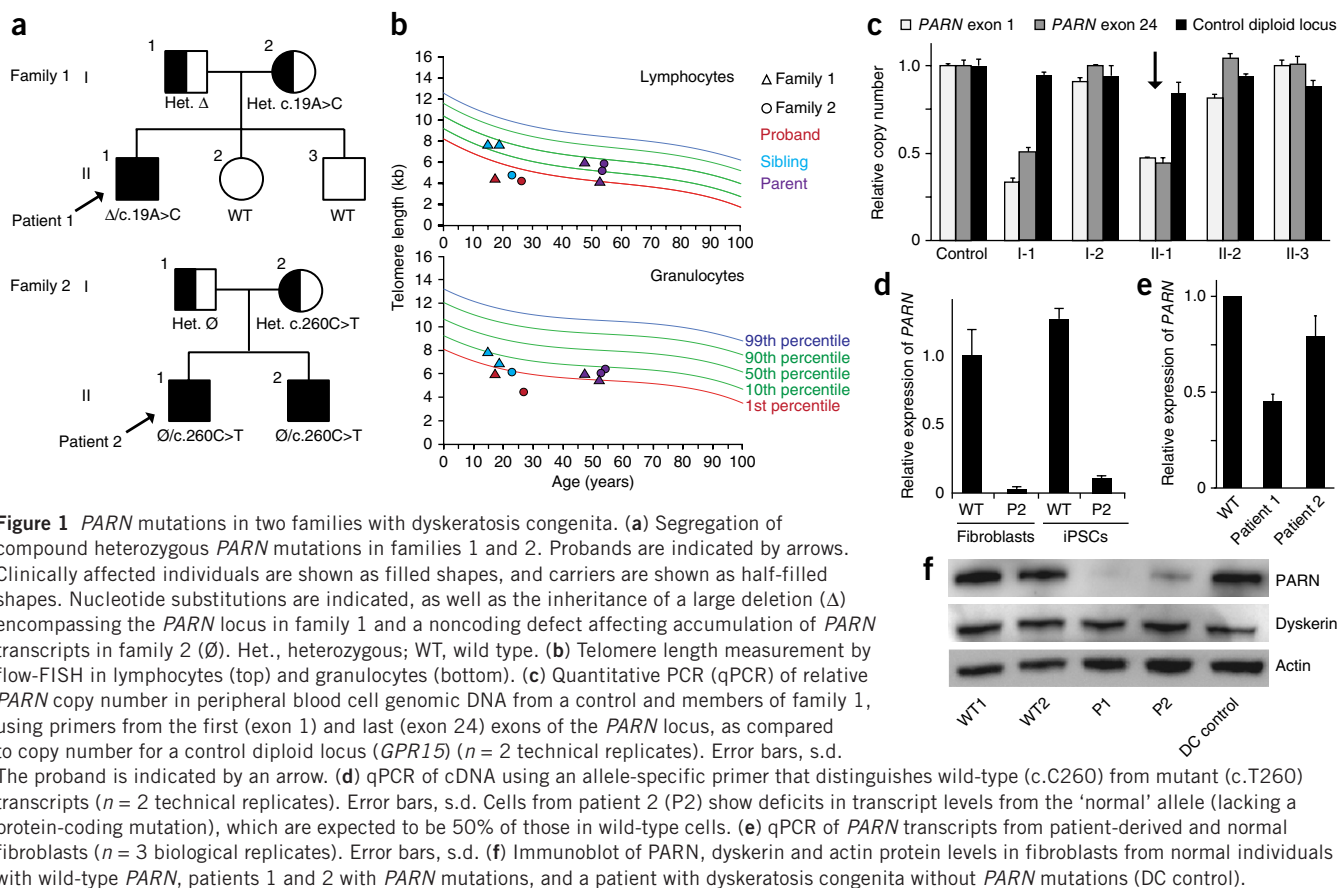
snoRNAs[8]. Here we show that cells from patients with *PARN* mutations manifest decreased levels of mature TERC RNA and increased levels of nascent transcripts with oligo(A) tails, which target nuclear RNAs for degradation[23,24]. Our results demonstrate a critical role for PARN in TERC biogenesis and provide a mechanism by which *PARN* mutations cause telomere disease.

By candidate gene sequencing in subjects with genetically uncharacterized dyskeratosis congenita or telomere disease in the Pediatric Myelodysplastic Syndrome and Bone Marrow Failure Registry at Boston Children's Hospital, we identified biallelic defects in *PARN* in two families (**Fig. 1a**), representing ~15% of the unrelated families in this disease category. The probands manifested classic features of dyskeratosis congenita and associated phenotypes, including bone marrow failure and very short telomere lengths in peripheral blood cells (**Fig. 1b** and **Supplementary Table 1**). Patient 1 was found to carry an undescribed missense variant, c.19A>C, on the allele inherited from his mother, resulting in the substitution of a highly conserved amino acid, p.Asn7His, and a large deletion encompassing the entire *PARN* gene on the allele inherited from his father (**Fig. 1c**, **Supplementary Fig. 1** and **Supplementary Table 2**). Patient 2 was found to carry a heterozygous missense variant, c.260C>T, encoding the substitution of a highly conserved amino acid, p.Ser87Leu, inherited from his unaffected mother (**Supplementary Fig. 2a,b** and **Supplementary Table 2**). He and his affected brother had no other potentially pathogenic exon-encoded variants. However, for the proband, we found decreased levels of *PARN* mRNA transcripts from the 'normal' allele, indicating a noncoding defect on the allele inherited from his father (**Fig. 1d** and **Supplementary Fig. 2c**). In keeping with these findings, *PARN* mRNA levels were diminished in fibroblasts from both patients (**Fig. 1e**), and PARN protein levels were even more severely compromised (**Fig. 1f**). These data show compound heterozygous loss-of-function mutations in *PARN* in two families with dyskeratosis congenita and very short telomeres.

We found that fibroblasts from patients with *PARN* alterations had decreased steady-state levels of *TERC* (**Fig. 2a**). Diminished *TERC* levels can result from disruption of dyskerin (encoded by *DKC1*),

Figure 1 *PARN* mutations in two families with dyskeratosis congenita. (**a**) Segregation of compound heterozygous *PARN* mutations in families 1 and 2. Probands are indicated by arrows. Clinically affected individuals are shown as filled shapes, and carriers are shown as half-filled shapes. Nucleotide substitutions are indicated, as well as the inheritance of a large deletion (Δ) encompassing the *PARN* locus in family 1 and a noncoding defect affecting accumulation of *PARN* transcripts in family 2 (Ø). Het., heterozygous; WT, wild type. (**b**) Telomere length measurement by flow-FISH in lymphocytes (top) and granulocytes (bottom). (**c**) Quantitative PCR (qPCR) of relative *PARN* copy number in peripheral blood cell genomic DNA from a control and members of family 1, using primers from the first (exon 1) and last (exon 24) exons of the *PARN* locus, as compared to copy number for a control diploid locus (*GPR15*) (*n* = 2 technical replicates). Error bars, s.d. The proband is indicated by an arrow. (**d**) qPCR of cDNA using an allele-specific primer that distinguishes wild-type (c.C260) from mutant (c.T260) transcripts (*n* = 2 technical replicates). Error bars, s.d. Cells from patient 2 (P2) show deficits in transcript levels from the 'normal' allele (lacking a protein-coding mutation), which are expected to be 50% of those in wild-type cells. (**e**) qPCR of *PARN* transcripts from patient-derived and normal fibroblasts (*n* = 3 biological replicates). Error bars, s.d. (**f**) Immunoblot of PARN, dyskerin and actin protein levels in fibroblasts from normal individuals with wild-type *PARN*, patients 1 and 2 with *PARN* mutations, and a patient with dyskeratosis congenita without *PARN* mutations (DC control).

which binds and stabilizes *TERC*[18], but cells from these patients did not carry mutations in *DKC1* or demonstrate reductions in dyskerin protein levels (**Fig. 1f**). To overcome limitations on cell numbers and study the effects of the *PARN* alterations in telomerase-expressing cells, we generated iPSCs from the fibroblasts of patients 1 and 2. These iPSCs also showed a deficiency in *PARN* transcripts and reduced *TERC* levels in comparison to normal iPSCs, without diminished dyskerin protein levels (**Fig. 2b–d**). We found that *PARN*-mutant iPSCs manifested decreased telomerase activity and impaired telomere elongation capacity in comparison to normal cells (**Fig. 2e,f**) but exhibit continuous self-renewal (with >25 passages thus far for all clones). To further investigate how PARN affects *TERC* levels, we knocked down *PARN* in HEK293 cells and again found reductions in *TERC* levels but not in the levels of *DKC1* mRNA or dyskerin protein (**Fig. 2g,h**). These studies demonstrate that PARN deficiency results in diminished *TERC* levels independently of *DKC1* (dyskerin), as well as in deficits in telomerase activity and telomere maintenance.

Box H/ACA snoRNAs are encoded in introns and undergo maturation by exonucleolytic processing of spliced intermediates[25]. Nascent snoRNAs are subject to oligoadenylation by TRF4-2 (refs. 8,26), a component of the TRAMP (TRF4-2, AIR2, MTR4) complex, which stimulates degradation of nuclear RNAs by the exosome[27]. PARN counteracts snoRNA oligoadenylation and promotes 3′-end maturation, which may prevent further TRAMP-mediated oligo(A) addition and subsequent degradation[8]. *TERC* is transcribed by RNA polymerase II (Pol II) as an autonomous genetic unit. The transcript has a 7-methylguanosine cap and a precise 3′ end, but, unlike other Pol II–dependent transcripts, *TERC* RNA does not contain a long poly(A) tail, and the mechanisms of its 3′-end maturation are

unknown[11,22,28,29]. On the basis of the box H/ACA architecture shared by snoRNAs and *TERC* RNA and the recent finding that *TERC* exists in oligo(A)-containing forms[30], we hypothesized that PARN participates in *TERC* 3′-end maturation and stabilization by removing oligo(A) tails from nascent transcripts.

We verified the presence of oligo(A)-containing forms of *TERC* by oligo(dT)$_{10}$ priming of total RNA for cDNA synthesis (**Fig. 3a**) and compared the relative proportions of oligo(A) and total *TERC*. We found a higher proportion of oligo(A) *TERC* in *PARN*-mutant fibroblasts and iPSCs than in normal cells (**Fig. 3b**). Because *TERC* oligo(A) tails average 6–7 nt in length[30] and may not be quantified accurately using oligo(dT)$_{10}$ priming, we performed 3′ RACE analysis (**Fig. 3c**). We observed an alteration in the size distribution of *TERC* 3′ ends in both patient-derived fibroblasts and iPSCs in comparison to normal cells, in that the intensity of the amplicon at the expected size was reduced and larger, extended transcripts predominated (**Fig. 3d**). These results were recapitulated in HEK293 cells upon transient *PARN* knockdown (**Fig. 3d**). By RNA blotting, using denaturing polyacrylamide gels, *in vitro*–transcribed *TERC* and endogenous mature TERC migrate as multiple bands[22]. In addition to these bands, we found more slowly migrating species in PARN-deficient cells, constituting a high proportion (~50%) of total *TERC*, in comparison to normal cells, and these species are likely to represent the extended *TERC* transcripts (**Fig. 3e,f**). These data indicate that PARN deficiency results not only in diminished *TERC* steady-state levels but also in increased levels of *TERC* RNA species with abnormal 3′ ends.

To investigate this finding more closely, we profiled the 3′ ends of *TERC* RNA in fibroblasts and iPSCs by deep sequencing (**Fig. 4**). We found a significant increase (*P* ≤ 0.05) in the proportion of *TERC*
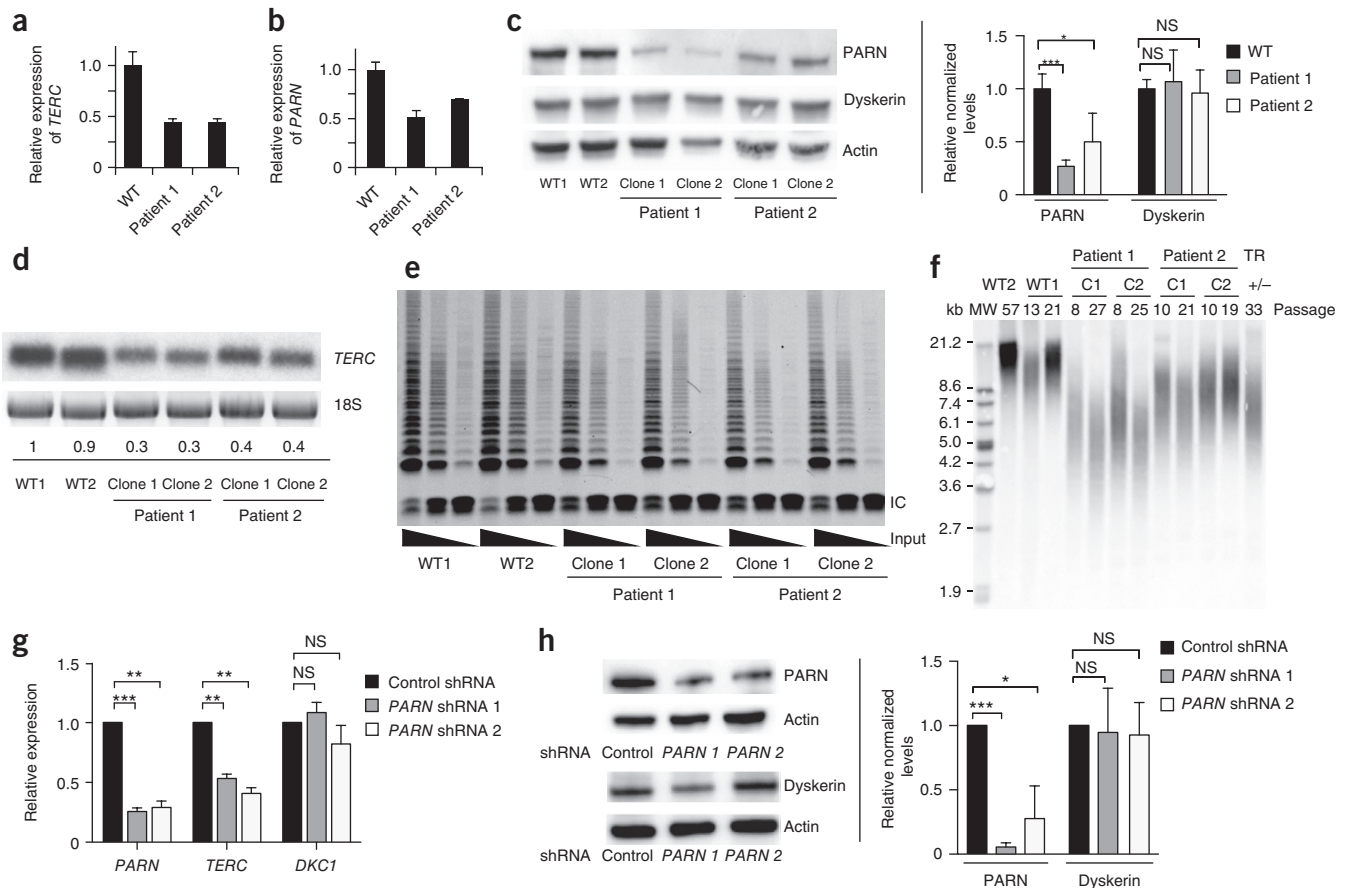
Figure 2 PARN deficiency results in decreased *TERC* levels, telomerase activity and telomere length. (**a**) qPCR of *TERC* transcripts in patient-derived and normal (wild-type) fibroblasts ($n = 2$ biological replicates). (**b**) qPCR of *PARN* transcript levels in iPSCs ($n = 2$ biological replicates). (**c**) Left, representative immunoblot of PARN, dyskerin and actin protein levels in iPSCs. Right, PARN and dyskerin protein levels normalized to actin levels ($n = 4$ biological replicates). (**d**) RNA blot of *TERC* following denaturing agarose gel electrophoresis of RNA from iPSCs. Ethidium bromide staining of 18S rRNA was used as a loading control. *TERC* levels normalized to those in WT1 cells are indicated. (**e**) Telomere repeat amplification protocol (TRAP) assay for telomerase activity in iPSCs, using fivefold dilutions of input cell extract. The internal control (IC) amplification standard is indicated. (**f**) Southern blot of telomere length by terminal restriction fragment length analysis in normal and patient-derived (clones C1 and C2) iPSCs, as well as control *TERC*-haploinsufficient (TR+/−) iPSCs. Passage numbers are indicated. MW, molecular weight marker. (**g**) qPCR of *PARN*, *TERC* and *DKC1* transcripts from HEK293 cells transduced with lentivirus encoding shRNA directed against *PARN* versus luciferase (control) ($n = 4$ biological replicates). (**h**) Left, representative immunoblot of PARN, dyskerin and actin protein levels in HEK293 cells (from **g**) transduced with lentivirus encoding shRNA directed against *PARN* versus luciferase. Right, PARN and dyskerin protein levels normalized to actin levels ($n = 4$ biological replicates). For all panels, error bars represent s.d. Significance is indicated: *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$; NS, not significant.

species other than mature TERC in patient-derived cells in comparison to normal cells, including transcripts extended beyond the canonical 3′ end and a diverse range of non-genomically encoded additions (**Supplementary Fig. 3**). We focused our attention on the distribution and frequency of two classes of *TERC* species: those containing up to eight genomically encoded bases beyond the canonical *TERC* end and the corresponding oligo(A) forms, which together comprised the majority of *TERC* forms across all samples. In patient-derived fibroblasts, we found that the percentage of mature TERC (ending at the canonical 5′-AUGC-3′ sequence) was markedly reduced in comparison to normal fibroblasts (59% versus 84%; **Fig. 4a**). Because the total amount of *TERC* in patient-derived fibroblasts was ~45% of that in normal fibroblasts (**Fig. 2a**), the level of mature TERC in patient-derived fibroblasts was approximately one-third that found in normal fibroblasts (27% versus 84%). At the same time, we found a significant increase in the proportion of oligo(A) forms of *TERC*, including transcripts with genomically encoded sequences beyond the 3′ end of mature TERC (**Fig. 4a**). Haploinsufficiency due to deletion

of one *PARN* allele in fibroblasts from the father of patient 1 yielded an intermediate phenotype (**Supplementary Fig. 4**). In iPSCs, mature TERC comprised ~70% of all species in normal cells but only ~40% in patient-derived cells (**Fig. 4b**). Again, we found significant increases in the proportion of oligo(A) forms of *TERC* transcripts extended beyond the 3′ terminus of mature TERC, with these transcripts, in aggregate, comprising the majority of the total *TERC* species in patient-derived iPSCs (**Fig. 3b**). Next, we profiled HEK293 cells in which *PARN* had been disrupted by RNA interference and similarly found diminished levels of mature TERC and an increased proportion of oligo(A) *TERC* forms (**Supplementary Fig. 5**). Sanger sequencing of 3′ RACE products corroborated the deep sequencing results (**Supplementary Fig. 6**). Consistently across all profiling experiments in patient-derived cells and cells with *PARN* knockdown, we observed a significant enrichment of oligo(A) forms of nascent *TERC* transcripts that were extended three, four or six bases beyond the canonical 3′ end. Collectively, the analyses performed in the setting of PARN deficiency show intermediates in TERC biogenesis

**Figure 3** PARN deficiency results in abnormal *TERC* 3′ ends.
(**a**) qPCR of *TERC* transcripts in cDNA from normal and patient-derived
fibroblasts generated using random hexamer versus oligo(dT)$_{10}$ priming
($n = 3$ technical replicates). Error bars, s.d. (**b**) Estimated proportion of
oligo(A)-containing *TERC* forms in fibroblasts and iPSCs ($n = 2$ technical
replicates). Error bars, s.d. (**c**) The 3′ RACE strategy. A universal
linker is ligated to the 3′ ends of total RNA. cDNA synthesis with a
linker-specific primer followed by PCR using linker- and *TERC*-specific
primers yields amplicons representing the diversity of *TERC* 3′ ends.
(**d**) 3′ RACE products from normal and patient-derived fibroblasts (left)
and iPSCs (middle) and from HEK293 cells subjected to shRNA-mediated
knockdown of *PARN* versus luciferase (control) (right). Amplicons were
separated by agarose gel electrophoresis; the expected size (184 bp)
of mature TERC is indicated (arrows). (**e**) RNA blot of *TERC* using RNA
from iPSCs, separated by denaturing 5% PAGE. U1 small nuclear RNA
(snRNA) represents loading control. The migration pattern of *in vitro*–
transcribed full-length *TERC* RNA is shown (IVT) and is considered to
represent mature TERC. Extended *TERC* species found in patient-derived
iPSCs are indicated on the right. Levels of total *TERC* species normalized
to the levels in WT1 are shown. The percentage of extended forms in
comparison to total *TERC* levels within each sample is indicated. Discrete,
extended TERC species could not be distinguished in wild-type iPSCs
(ND). (**f**) RNA blot (urea 5% PAGE) of *TERC* using RNA from HEK293
cells subjected to shRNA-mediated knockdown of *PARN* (sh) versus
luciferase. Analysis was as described in **e**.
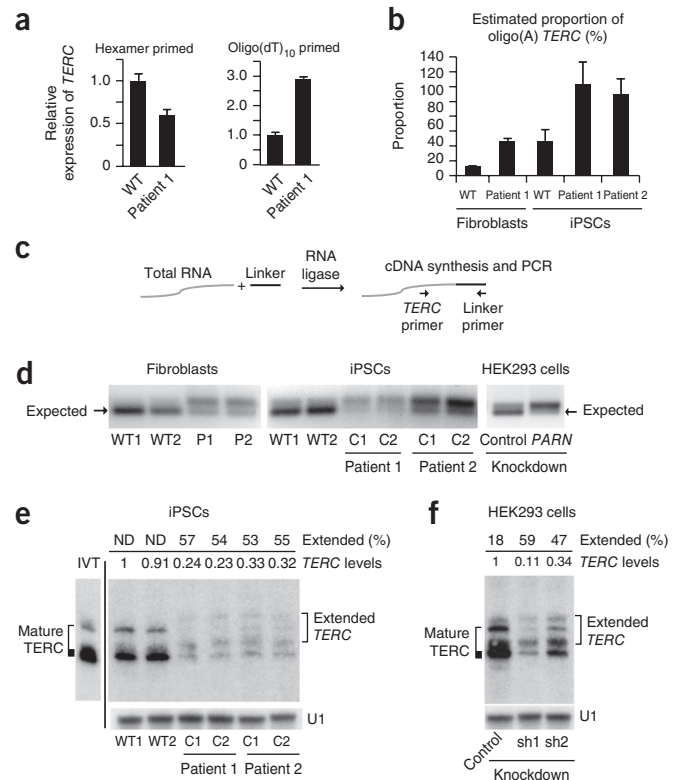
and indicate a critical and non-redundant function for PARN in
counteracting the oligoadenylation of nascent *TERC* transcripts.

Oligoadenylation is predicted to target nuclear RNAs for degra-
dation, and we therefore determined the decay rate of *TERC* RNA
in PARN-deficient versus normal cells. We inhibited Pol II trans-
cription using actinomycin D and quantified *TERC* RNA levels
by RNA blotting. We found a decreased half-life for *TERC* tran-
scripts in patient-derived iPSCs in comparison to normal iPSCs and
also in HEK293 cells with *PARN* knockdown in comparison to

cells with control knockdown (**Fig. 5**). These results indicate
that PARN deficiency results in destabilization of *TERC* RNA
and provide a mechanism for the observed decrease in steady-state
TERC levels.
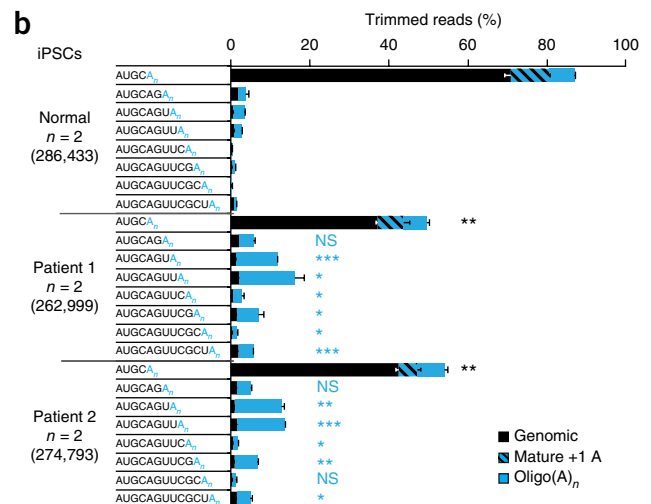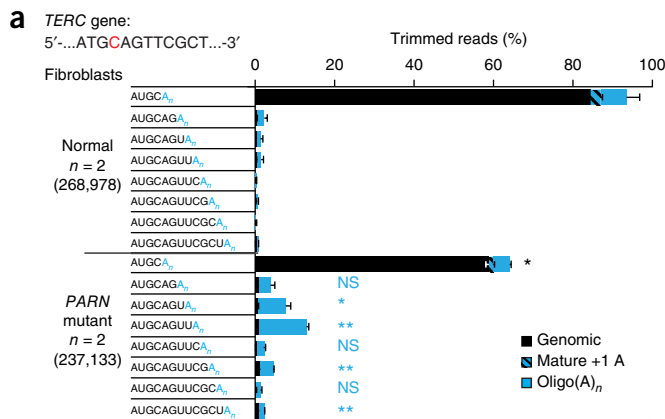
Figure 4 Decreased proportion of mature TERC and increased proportion of 3′-extended *TERC* species in PARN-deficient patient-derived cells.
(**a**) Fibroblasts. 3′ RACE PCR products from normal and *PARN*-mutant patient-derived fibroblasts were subjected to deep sequencing, and reads were
aligned to the *TERC* gene. The canonical TERC 3′ terminus is shown in red in the context of the *TERC* gene sequence. The relative proportions of mature
TERC and *TERC* RNA species extending up to eight bases into the genomic sequence, with or without post-transcriptionally added oligo(A) tails, are
depicted. Species with genomically encoded termini are in black, mature TERC with a single adenosine (which may be genomically encoded) is hatched
and species with oligo(A) additions of any length (n) are in solid blue. Proportions are averaged for normal fibroblasts and patient-derived fibroblasts ($n = 2$
for each group). The total number of trimmed reads for each group is shown in parentheses. Error bars, s.d. For statistical evaluations, mature TERC forms
and all genomically extended *TERC* species with oligo(A)$_n$ ends were compared between normal and patient-derived cells in a two-tailed *t* test. Significance
is indicated: $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$; NS, not significant (black, mature TERC; blue, extended, oligo(A)$_n$ forms of *TERC*). (**b**) iPSCs.
3′ RACE PCR products from normal and *PARN*-mutant patient-derived iPSCs were subjected to deep sequencing, and reads were aligned to the *TERC* gene.
Analyses and statistical comparisons were performed as in **a**. Proportions are averaged for normal iPSCs and patient-derived iPSCs ($n = 2$ for each group).

Figure 5 Decreased stability of *TERC* in PARN-deficient cells. (**a**) Representative RNA blot of *TERC* RNA levels in normal versus patient-derived iPSCs, at 0, 1, 2 and 4 h after the addition of actinomycin D. Results for WT1 and patient 2 clone 2 iPSCs are shown. Ethidium bromide staining of 28S and 18S rRNAs is shown; 18S rRNA was used as a loading control. *TERC* levels normalized to those at time 0 for each sample are indicated. (**b**) Graph of the *TERC* decay rate in normal and patient-derived iPSCs, calculated from RNA blot analysis as shown in **a** (*n* = 2 for each group; error bars, s.d.). The dotted lines reflect the slope determined by simple linear regression, and half-life is the calculated *x* intercept at *y* = 0.5. (**c,d**) RNA blot (**c**) and decay rates (**d**) of *TERC* RNA in HEK293 cells subjected to shRNA-mediated knockdown of *PARN* versus control. Quantification was performed as in **a** and **b**.
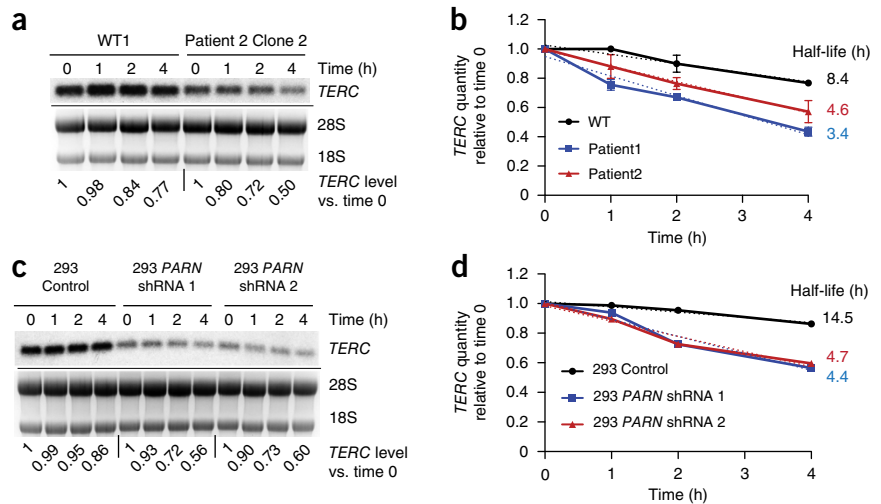
Because the canonical role of PARN is believed to be in mRNA turnover, we performed RNA sequencing (RNA-seq) to assess the effects of PARN deficiency on mRNAs globally. Remarkably, we found that no protein-coding mRNAs manifested a fold change (increase or decrease) in steady-state levels that consistently exceeded the change in TERC levels in all seven sample pairs (**Supplementary Tables 3 and 4**). These results suggest that the degree of PARN deficiency that is sufficient to disrupt TERC RNA transcript levels, 3′-end processing
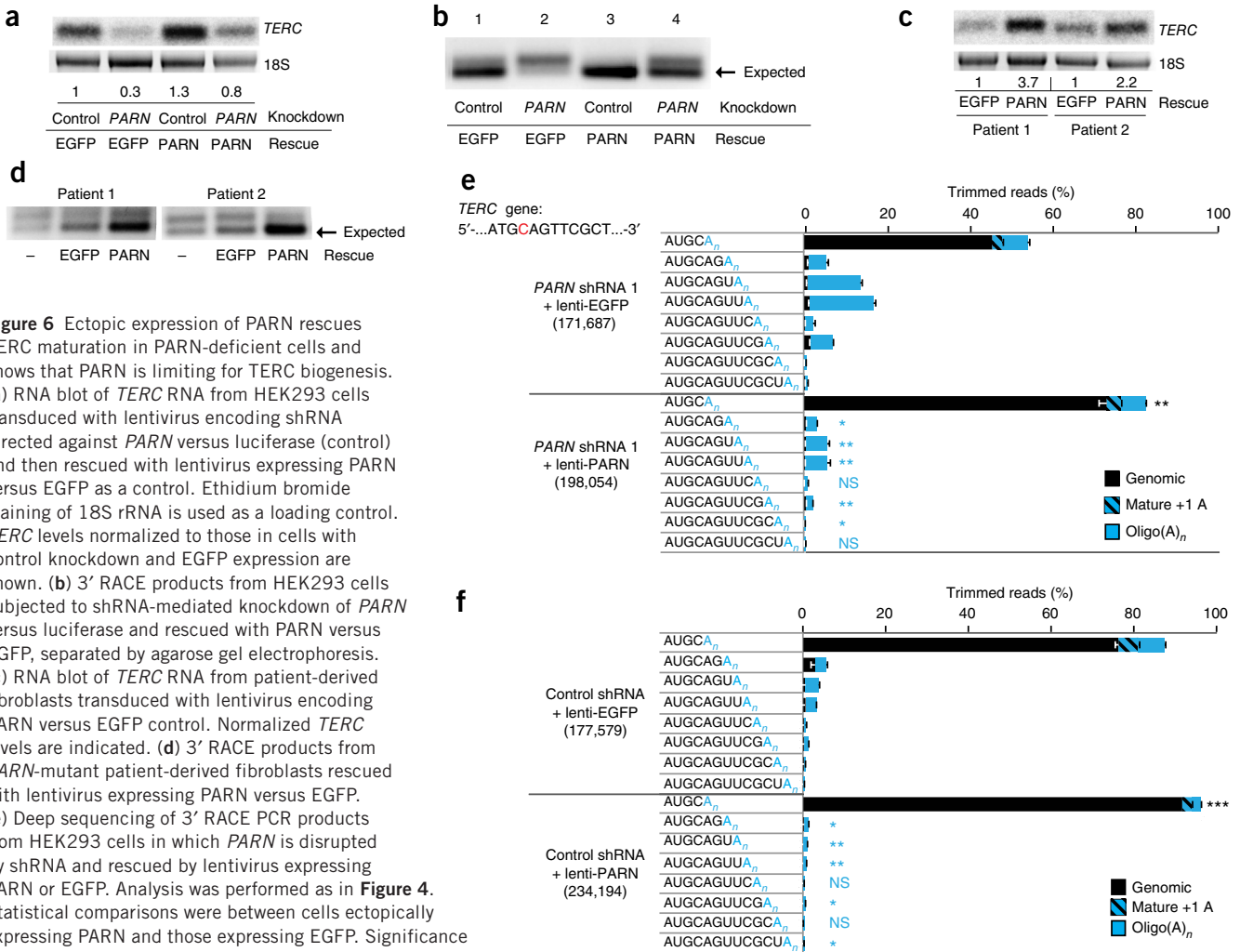
Figure 6 Ectopic expression of PARN rescues TERC maturation in PARN-deficient cells and shows that PARN is limiting for TERC biogenesis. (**a**) RNA blot of *TERC* RNA from HEK293 cells transduced with lentivirus encoding shRNA directed against *PARN* versus luciferase (control) and then rescued with lentivirus expressing PARN versus EGFP as a control. Ethidium bromide staining of 18S rRNA is used as a loading control. *TERC* levels normalized to those in cells with control knockdown and EGFP expression are shown. (**b**) 3′ RACE products from HEK293 cells subjected to shRNA-mediated knockdown of *PARN* versus luciferase and rescued with PARN versus EGFP, separated by agarose gel electrophoresis. (**c**) RNA blot of *TERC* RNA from patient-derived fibroblasts transduced with lentivirus encoding PARN versus EGFP control. Normalized *TERC* levels are indicated. (**d**) 3′ RACE products from *PARN*-mutant patient-derived fibroblasts rescued with lentivirus expressing PARN versus EGFP. (**e**) Deep sequencing of 3′ RACE PCR products from HEK293 cells in which *PARN* is disrupted by shRNA and rescued by lentivirus expressing PARN or EGFP. Analysis was performed as in **Figure 4**. Statistical comparisons were between cells ectopically expressing PARN and those expressing EGFP. Significance is indicated: *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$; NS, not significant (black, mature TERC; blue, extended, oligo(A)$_n$ *TERC* forms). Error bars, s.d. (**f**) Deep sequencing of 3′ RACE PCR products from HEK293 control cells transduced with lentivirus overexpressing PARN versus EGFP. Statistical comparisons were performed as in **e**. Error bars, s.d.

and stability does not result in changes of a similar magnitude in the levels of mRNAs expressed across cell types.

We next investigated the effects of ectopic PARN expression on *TERC* 3′-end processing. We used lentiviral vectors to overexpress PARN in HEK293 cells where *PARN* had been knocked down using a hairpin (shRNA 1) targeting the 3′ UTR (**Supplementary Fig. 7**). We found not only rescue of the diminished steady-state levels of *TERC* (**Fig. 6a**) but also restoration of *TERC* processing, as evidenced by an increased proportion of 3′ RACE amplicons of the expected size relative to extended forms (**Fig. 6b**). Similar results were obtained in patient-derived fibroblasts (**Fig. 6c,d**). Deep sequencing confirmed an increased proportion of mature TERC (72% versus 45%) and a decreased proportion of oligo(A) *TERC* forms after PARN rescue in HEK293 cells (**Fig. 6e**) and partial restoration of 3′-end processing with ectopic PARN expression in patient-derived fibroblasts (**Supplementary Fig. 8**). We were unable to increase the levels of PARN protein in patient-derived iPSCs using lentivirus vectors, possibly because of silencing or toxicity (**Supplementary Fig. 9**). Notably, when we overexpressed PARN in control HEK293 cells (**Supplementary Fig. 7b**), we found a significant increase in the proportion of mature TERC (92% versus 75% in control cells) as well as a decrease in the proportion of extended species and oligo(A) forms (**Fig. 6f**). Taken together, these studies demonstrate that the defects in *TERC* biogenesis in patients with *PARN* alterations can be restored by ectopic expression of PARN. Moreover, PARN appears to be limiting for *TERC* post-transcriptional processing and thus may have a pivotal role in determining telomerase levels in cells.

It is well established that the level of TERC in cells is important in health and disease[13], including in tissue renewal[31–35], degenerative disorders[20,36–41], cancer[15,16,42] and possibly longevity[43,44], but the mechanisms by which TERC levels are regulated have remained elusive. Our findings, emerging from genetic discovery in patients with telomere diseases[1,2], demonstrate that PARN mediates the post-transcriptional maturation of *TERC*. We propose that PARN removes oligo(A) tails and/or genomically encoded terminal nucleotides from nascent *TERC* transcripts, leading to maturation of the 3′ end and protection from further TRF4-2–mediated oligoadenylation and degradation by the exosome (**Supplementary Fig. 10**). We cannot exclude the possibility that *PARN* mutations contribute to disease manifestations by influencing the levels of other RNAs. However, given the phenotypic similarity of patients with *PARN* alterations and other forms of TERC deficiency, our results suggest that disrupting the role of PARN in TERC biogenesis may be sufficient to cause dyskeratosis congenita or telomere diseases. On the basis of our findings and those of others investigating the global effects of *PARN* disruption[8,45], we speculate that a major, non-redundant role of PARN in mammalian cells is not in mRNA metabolism but in regulating the biogenesis of TERC and other noncoding RNAs.

URLs. Pediatric Myelodysplastic Syndrome and Bone Marrow Failure Registry, http://www.pedimds.org/; Ensembl data download, ftp://ftp.ensembl.org/pub/release-80/fasta/homo_sapiens/.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. RNA-seq data have been deposited in the Gene Expression Omnibus (GEO) under accession GSE71709.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

1. Stuart, B.D. *et al.* Exome sequencing links mutations in *PARN* and *RTEL1* with familial pulmonary fibrosis and telomere shortening. *Nat. Genet.* **47**, 512–517 (2015).
2. Tummala, H. *et al.* Poly(A)-specific ribonuclease deficiency impacts telomere biology and causes dyskeratosis congenita. *J. Clin. Invest.* **125**, 2151–2160 (2015).
3. Dehlin, E., Wormington, M., Korner, C.G. & Wahle, E. Cap-dependent deadenylation of mRNA. *EMBO J.* **19**, 1079–1086 (2000).
4. Körner, C.G. & Wahle, E. Poly(A) tail shortening by a mammalian poly(A)-specific 3′-exoribonuclease. *J. Biol. Chem.* **272**, 10448–10456 (1997).
5. Körner, C.G. *et al.* The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of *Xenopus* oocytes. *EMBO J.* **17**, 5427–5437 (1998).
6. Virtanen, A., Henriksson, N., Nilsson, P. & Nissbeck, M. Poly(A)-specific ribonuclease (PARN): an allosterically regulated, processive and mRNA cap–interacting deadenylase. *Crit. Rev. Biochem. Mol. Biol.* **48**, 192–209 (2013).
7. Yoda, M. *et al.* Poly(A)-specific ribonuclease mediates 3′-end trimming of Argonaute2-cleaved precursor microRNAs. *Cell Rep.* **5**, 715–726 (2013).
8. Berndt, H. *et al.* Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* **18**, 958–972 (2012).
9. Mason, P.J. & Bessler, M. mRNA deadenylation and telomere disease. *J. Clin. Invest.* **125**, 1796–1798 (2015).
10. Egan, E.D. & Collins, K. An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. *Mol. Cell. Biol.* **32**, 2428–2439 (2012).
11. Feng, J. *et al.* The RNA component of human telomerase. *Science* **269**, 1236–1241 (1995).
12. Venteicher, A.S. *et al.* A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644–648 (2009).
13. Greider, C.W. Telomerase RNA levels limit the telomere length equilibrium. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 225–229 (2006).
14. Cristofari, G. & Lingner, J. Telomere length homeostasis requires that telomerase levels are limiting. *EMBO J.* **25**, 565–574 (2006).
15. Cao, Y., Bryan, T.M. & Reddel, R.R. Increased copy number of the *TERT* and *TERC* telomerase subunit genes in cancer cells. *Cancer Sci.* **99**, 1092–1099 (2008).
16. Soder, A.I. *et al.* Amplification, increased dosage and *in situ* expression of the telomerase RNA gene in human cancer. *Oncogene* **14**, 1013–1021 (1997).
17. Heiss, N.S. *et al.* X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat. Genet.* **19**, 32–38 (1998).
18. Mitchell, J.R., Wood, E. & Collins, K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402**, 551–555 (1999).
19. Vulliamy, T. *et al.* Mutations in the telomerase component *NHP2* cause the premature ageing syndrome dyskeratosis congenita. *Proc. Natl. Acad. Sci. USA* **105**, 8073–8078 (2008).
20. Vulliamy, T. *et al.* The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* **413**, 432–435 (2001).
21. Walne, A.J. *et al.* Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. *Hum. Mol. Genet.* **16**, 1619–1629 (2007).
22. Mitchell, J.R., Cheng, J. & Collins, K. A box H/ACA small nucleolar RNA–like domain at the human telomerase RNA 3′ end. *Mol. Cell. Biol.* **19**, 567–576 (1999).
23. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).
24. Schmidt, K. & Butler, J.S. Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity. *Wiley Interdiscip. Rev. RNA* **4**, 217–231 (2013).
25. Kiss, T., Fayet-Lebaron, E. & Jady, B.E. Box H/ACA small ribonucleoproteins. *Mol. Cell* **37**, 597–606 (2010).
26. Rammelt, C., Bilen, B., Zavolan, M. & Keller, W. PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA* **17**, 1737–1746 (2011).

27. LaCava, J. *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713–724 (2005).
28. Fu, D. & Collins, K. Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs. *Mol. Cell* **11**, 1361–1372 (2003).
29. Zaug, A.J., Linger, J. & Cech, T.R. Method for determining RNA 3′ ends and application to human telomerase RNA. *Nucleic Acids Res.* **24**, 532–533 (1996).
30. Goldfarb, K.C. & Cech, T.R. 3′ terminal diversity of MRP RNA and other human noncoding RNAs revealed by deep sequencing. *BMC Mol. Biol.* **14**, 23 (2013).
31. Jongmans, M.C. *et al.* Revertant somatic mosaicism by mitotic recombination in dyskeratosis congenita. *Am. J. Hum. Genet.* **90**, 426–433 (2012).
32. Kirwan, M. *et al.* Exogenous TERC alone can enhance proliferative potential, telomerase activity and telomere length in lymphocytes from dyskeratosis congenita patients. *Br. J. Haematol.* **144**, 771–781 (2009).
33. Westin, E.R. *et al.* Telomere restoration and extension of proliferative lifespan in dyskeratosis congenita fibroblasts. *Aging Cell* **6**, 383–394 (2007).
34. Wong, J.M. & Collins, K. Telomerase RNA level limits telomere maintenance in X-linked dyskeratosis congenita. *Genes Dev.* **20**, 2848–2858 (2006).
35. Hao, L.Y. *et al.* Short telomeres, even in the presence of telomerase, limit tissue renewal capacity. *Cell* **123**, 1121–1131 (2005).
36. Alter, B.P., Giri, N., Savage, S.A. & Rosenberg, P.S. Cancer in dyskeratosis congenita. *Blood* **113**, 6549–6557 (2009).
37. Armanios, M.Y. *et al.* Telomerase mutations in families with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* **356**, 1317–1326 (2007).
38. Calado, R.T. *et al.* Constitutional telomerase mutations are genetic risk factors for cirrhosis. *Hepatology* **53**, 1600–1607 (2011).
39. Kirwan, M. *et al.* Defining the pathogenic role of telomerase mutations in myelodysplastic syndrome and acute myeloid leukemia. *Hum. Mutat.* **30**, 1567–1573 (2009).
40. Tsakiri, K.D. *et al.* Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc. Natl. Acad. Sci. USA* **104**, 7552–7557 (2007).
41. Yamaguchi, H. *et al.* Mutations of the human telomerase RNA gene (*TERC*) in aplastic anemia and myelodysplastic syndrome. *Blood* **102**, 916–918 (2003).
42. Trapp, S. *et al.* A virus-encoded telomerase RNA promotes malignant T cell lymphomagenesis. *J. Exp. Med.* **203**, 1307–1317 (2006).
43. Codd, V. *et al.* Common variants near *TERC* are associated with mean telomere length. *Nat. Genet.* **42**, 197–199 (2010).
44. Soerensen, M. *et al.* Genetic variation in *TERT* and *TERC* and human leukocyte telomere length and longevity: a cross-sectional and longitudinal analysis. *Aging Cell* **11**, 223–227 (2012).
45. Lee, J.E. *et al.* The PARN deadenylase targets a discrete set of mRNAs for decay and regulates cell motility in mouse myoblasts. *PLoS Genet.* **8**, e1002901 (2012).

ONLINE METHODS

Patient material. The probands and families were enrolled in the Pediatric Myelodysplastic Syndrome and Bone Marrow Failure Registry at Boston Children's Hospital. Biological samples were procured under protocols approved by the Institutional Review Board at Boston Children's Hospital after obtaining written informed consent in accordance with the Declaration of Helsinki.

Telomere length measurements. *Flow-FISH.* Telomere length in peripheral blood cell subsets was measured using flow-FISH[46] by Repeat Diagnostics.

Terminal restriction fragment length analysis. Genomic DNA (1.8 µg) was digested with the HinfI and RsaI restriction enzymes and separated on 0.8% agarose gels, followed by Southern blot analysis using the TeloTAGGG Telomere Length Assay kit (Roche).

Primary cells and cell lines. Fibroblasts were cultured from skin biopsies obtained from patients and healthy volunteer subjects under approved protocols. Briefly, 2-mm punch biopsies were diced and placed under a coverslip in DMEM supplemented with 15% FCS until keratinocyte and fibroblast outgrowths were apparent. Fibroblasts were subcultured and expanded using 0.05% trypsin and DMEM supplemented with 15% FCS. The normal skin fibroblasts used in these experiments were from healthy adult volunteers and included NHSF2 (a gift from A. Klingelhutz, University of Iowa) and Hfib2 (ref. 47) cells. HEK293 and HEK293T cells (obtained from the American Type Culture Collection) were subcultured and expanded using 0.05% trypsin and DMEM supplemented with 10% FCS. Derivation, characterization and culture of iPSCs from fibroblasts for patient 1 were performed as described[48]. NHSF2 cells and fibroblasts from patient 2 were reprogrammed using the pRRL.PPT. SF.hOKSMco.idTomato.preFRT lentiviral reprogramming vector[49] (a gift from A. Schambach, Hannover Medical School). WT1 iPSCs were derived from NHSF2 cells, and WT2 iPSCs were derived from hFib2 cells[47]. For feeder-free culture, iPSCs were maintained in Essential 8 medium (Life Technologies) on Matrigel matrix qualified for use with human embryonic stem cells (BD Biosciences). Cell lines were not tested for mycoplasma.

DNA sequencing and genetic analysis. *DNA isolation and sequencing.* Genomic DNA was extracted from primary cells and cell lines using the Gentra PureGene kit (Qiagen). DNA from research subject saliva samples was extracted with the prepIT kit (Oragene DNA Genotek). The primer sequences for *PARN* gene amplification and sequencing are provided in **Supplementary Table 5**. Sanger sequencing was performed by Genewiz.

Copy number determination by quantitative PCR. Copy number at the *PARN* locus in genomic DNA from fibroblasts, peripheral blood cells or saliva samples for family 1 was determined by performing qPCR using primers spanning *PARN* exon 1, *PARN* exon 24 and a control diploid locus (*GPR15*) (with PARNcopy_ex1L/R, PARNcopy_ex24L/R and GPR15copy_L/R primers, respectively; **Supplementary Table 6**). A standard curve was generated for each primer pair and used to calculate the relative copy number of each *PARN* amplicon, calibrated to *GPR15* copy number. Quantitative measurements were normalized to those obtained with genomic DNA from a healthy volunteer.

Genome-wide SNP microarray analysis. Microarray analysis was performed on peripheral blood DNA from patient 1 using the Infinium Assay with the Illumina CytoSNP-850K BeadChip platform by the Cytogenetics program at Cincinnati Children's Hospital Medical Center.

RNA isolation, cDNA synthesis and quantitative RT-PCR. RNA from patient blood samples was processed using the PAXgene Blood RNA kit (Qiagen). RNA from cultured cells was recovered using TRIzol (Ambion). After DNase treatment (Turbo DNA-free, Ambion), cDNA was synthesized using 1 µg of total RNA, 50 ng of random hexamers or 7.5 ng of oligo(dT)$_{10}$, and 1 µl of SuperScript III reverse transcriptase (Invitrogen) in a total volume of 20 µl for 1 h at 50 °C. qPCR was performed using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) and the primers listed in **Supplementary Table 6** (PARN_L/R, TERC_L1/R1, DKC1_L/R, POLR2A_L/R and ACTB_L/R) in a CFX96 Real-Time PCR detection system (Bio-Rad). *PARN*, *TERC* and *DKC1* levels were normalized to those for *POLR2A* and/or *ACTB*, which in direct comparisons gave similar results. Graphing and statistical analysis of qPCR

results were performed using GraphPad Prism. The sequences of the primers used for qPCR specific to the wild-type allele of the *PARN* gene in patient 2 (PARNcDNA_ex3L and PARNcDNA_ex5RWT) and for sequencing of *PARN* transcripts from peripheral blood cell cDNA from family 2 (PARNcDNA_ex4L and PARNcDNA_ex6R) are provided in **Supplementary Table 6**.

To estimate the proportion of *TERC* that was oligoadenylated relative to total *TERC* (**Fig. 3b**), standard curves were generated for *TERC* amplification using hexamer-primed cDNA and oligo(dT)$_{10}$-primed cDNA, after normalizing to *ACTB* levels. Oligo(A) *TERC* levels were derived using oligo(dT)$_{10}$-primed cDNA and represented as a percentage of total *TERC* levels, which were derived using hexamer-primed cDNA. The primers used were TERC_L/R and ACTB_L/R (**Supplementary Table 6**).

RNA blots. *Agarose-formaldehyde gel electrophoresis.* Total RNA (4–10 µg) was separated on 1.5% agarose-formaldehyde gels, transferred to HyBond N+ membranes (Amersham) in 10× SSC buffer by capillary transfer and hybridized with [α-^{32}P]dCTP-labeled full-length *TERC* probe in ULTRAhyb buffer (Life Technologies). Signals were normalized to 18S rRNA levels determined on the basis of ethidium bromide staining. Quantification was performed using ImageJ software.

Denaturing PAGE. Total RNA (1–2 µg) was separated on 5% polyacrylamide TBE-urea gels, transferred to HyBond N+ membranes by electroblotting and hybridized with [α-^{32}P]dCTP-labeled full-length *TERC* or U1 snRNA probe. *In vitro*–transcribed *TERC* RNA (20 pg) was loaded as a control. *TERC* signal was normalized to U1 snRNA signal. Quantification was performed using ImageJ software. The primers used to amplify the probes were TERC_L4 and TERC_R2 for *TERC* and U1snRNA_L and U1snRNA_R for U1 snRNA (**Supplementary Table 6**).

In vitro transcription of *TERC* RNA. The full-length *TERC* transcripts used for denaturing PAGE and RNA blots were generated by *in vitro* transcription using a T7 promoter–*TERC* PCR amplicon template (MAXIscript T7, Life Technologies). The primers used to generate the T7 promoter–*TERC* template were T7_TERC_L and TERC_R2 (**Supplementary Table 6**).

TERC RNA decay. iPSCs or HEK293 cells (0.5–1 × 10^6) were treated with 5 µg/ml actinomycin D (Life Technologies) and collected in TRIzol at 0, 1, 2 and 4 h after treatment. Purified RNA was subjected to RNA blotting by 1.5% agarose-formaldehyde gel electrophoresis, transfer to nylon membranes and hybridization as described above. *TERC* signal was normalized to 18S rRNA signal. The slopes for decay plots were determined by simple linear regression, and transcript half-life was calculated as the *x* intercept at *y* = 0.5, using GraphPad Prism.

Telomerase activity assays. iPSCs (2 × 10^5) were lysed in TRAPeze 1× CHAPS lysis buffer (Millipore), and protein was quantified by Bradford assay (Bio-Rad). Fivefold dilutions of cell extracts (containing 100 ng, 20 ng and 4 ng of protein) were subjected to the TRAP assays using the TRAPeze Telomerase Detection kit (Millipore). Products were resolved on 10% TBE polyacrylamide gels and visualized by staining with GelRed (Biotium).

3′ RACE. Total RNA (600 ng) was ligated to 5 µM of 5′-adenylated, 3′-blocked adaptor (Universal miRNA Cloning Linker, New England BioLabs) with 280 units of T4 RNA ligase, Truncated KQ (New England BioLabs), 25% PEG 8000 and 1 µl of RNaseOUT (Life Technologies) in a 20-µl reaction at 25 °C for 16–24 h. After cleanup with RNA Clean and Concentrator columns (Zymo Research), followed by DNase treatment, cDNA was synthesized with 5 pmol of universal RT primer (**Supplementary Table 6**) and SuperScript III reverse transcriptase. PCR amplification was carried out using 5 µM of the TERC_L2 and universal RT or TERC_L3 and universal RT primer sets (**Supplementary Table 6**) with SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). PCR products were directly analyzed on 2.5% agarose gels to visualize mature TERC and extended *TERC* transcripts or subjected to QIAquick PCR purification columns (Qiagen) for library preparation for deep sequencing. For Sanger sequencing, 3′ RACE PCR products were directly cloned into the pCR4_TOPO vector (Life Technologies), and individual clones were sequenced using the TERC_L2 or TERC_L3 primer.

MiSeq library preparation and analysis. RACE products were prepared for deep sequencing using the TruSeq Nano DNA LT Library Prep kit (Illumina). Briefly, for each sample, linkers carrying unique barcodes were ligated to the RACE products with DNA ligase. Ligated products were then amplified using primers containing Illumina adaptors and size selected using magnetic beads (AMPure). The completed libraries were submitted to the Tufts University Genomics Core for sequencing and data analysis. The quality and quantity of each library was determined on an Advanced Analytical Fragment Analyzer. The libraries were then pooled to equimolar concentrations. The pooled library was sequenced on an Illumina MiSeq instrument with paired-end 250-base reads using the Illumina TruSeq v2 500 Cycles kit. Reads were demultiplexed with CASAVA 1.8.2, and the read 1 and read 2 fastq files for each sample were generated. For data analysis, the Illumina adaptor sequence was first removed from the end of each read using Trimmomatic[50]. The resulting unbroken pair was then joined using FLASH[51]. Joined reads from each sample were mapped to the *TERC* gene (NR_001566) with the Bowtie 2 mapper[52] and outputted to a SAM file. The SAM file from each sample was then used as input for custom-developed Perl scripts to remove the 3′ RACE universal adaptor sequence, determine the position at which the genomic sequence ended, and determine the position and length of the oligo(A) tail.

Code availability: The Perl scripts for *TERC* RNA end analysis will be made available upon request.

RNA-seq library preparation and analysis. RNA integrity was verified using the Advanced Analytical Fragment Analyzer. Total RNA (0.5–1 μg) was used as input for library preparation with the TruSeq Stranded Total RNA with RiboZero Gold kit (Illumina). The molar concentrations of the libraries were determined using the Advanced Analytical Fragment Analyzer, and the libraries were pooled at equimolar concentrations. The pooled libraries were sequenced on two lanes of a HiSeq 2500 instrument with High-Output single-read 50-base format. Sequence reads were aligned to the human transcriptome using HISAT version 0.1.6 (ref. 53). The transcriptome file consisted of protein-coding and noncoding RNA sequences downloaded from Ensembl (release 80). To estimate transcript abundances, we applied Salmon version 0.3.2 (ref. 54) to the aligned reads and summarized transcript abundances into gene-level expression data by summing the expression levels for all transcripts mapping to the same gene. Gene-to-transcript mappings and transcript type annotations (for example, assignment of transcripts to categories such as snRNA, snoRNA, etc.) were also downloaded from Ensembl. Automated annotation was manually adjusted by annotating *TERC* and other small Cajal body–specific RNAs (scaRNAs) as snoRNAs for the purpose of this analysis.

We took two approaches to find genes that were commonly differentially expressed between normal and PARN-deficient cell lines. First, we performed the Wilcoxon signed-rank test on the 13,508 genes that had expression levels of at least 1 transcript per million (TPM) in one sample, using the following pairs of samples: WT1 fibroblasts versus patient 1 fibroblasts; WT1 IPSCs versus patient 1 clone 2 iPSCs; WT2 iPSCs versus patient 2 clone 1 iPSCs; and HEK293 cells with control knockdown versus HEK293 cells with *PARN* knockdown. No genes were differentially expressed with nominal significance ($P \leq 0.05$) across all pairs using this approach or using the paired *t* test. Second, because we showed that *TERC* levels are decreased in PARN-deficient cell lines by qRT-PCR and RNA blotting, we reasoned that the magnitude of the fold change in its levels would be an appropriate threshold to define other genes as differentially expressed in paired comparisons. The comparison-specific thresholds in absolute values, which we refer to as *TERC*-defined thresholds, are provided in **Supplementary Table 7**. We found that only one transcript, for a snoRNA (*SCARNA13*), showed a fold change in all seven comparisons that exceeded that for *TERC* (**Supplementary Table 4**). We next lessened the stringency of our ad hoc metric of differential expression by allowing genes to be considered differentially expressed if they exceeded *TERC*-defined thresholds in fewer (five or six) than all seven pairwise comparisons. In this analysis,

we asked whether the number of genes in each transcript category was different than would be expected by chance given the total number of genes differentially expressed, using the χ^2 test (**Supplementary Table 3**).

Immunoblot analysis. Total cellular lysates were subjected to SDS-PAGE, and protein was transferred to PVDF membranes using standard procedures. Detection of PARN and dyskerin was performed using primary antibodies to human PARN (Abcam, ab188333; 1:5,000 dilution) and dyskerin (Santa Cruz Biotechnology, sc-48794; 1:1,000 dilution) and horseradish peroxidase (HRP)-conjugated goat anti-rabbit secondary antibody (Bio-Rad, 170-5046; 1:15,000 dilution), followed by chemiluminescent detection using Clarity Western ECL substrate (Bio-Rad). Protein loading was determined using HRP-conjugated antibody to actin (Santa Cruz Biotechnology, sc-1615; 1:1,000 dilution) on the same membranes and used for normalization. Imaging and quantification of chemiluminescent signals was performed using the Bio-Rad ChemiDoc Touch imaging system. Graphing and statistical analyses were performed using GraphPad Prism.

Lentiviral RNA interference and cDNA expression. *shRNA constructs.* Duplex oligonucleotides encoding shRNAs targeting human *PARN* (NM_002582) or luciferase control (**Supplementary Table 6**) were cloned into the pLKO.1-puro vector (Addgene, 10878), a gift from D. Root (Broad Institute).

cDNA expression constructs. cDNAs encoding the *PARN* ORF (NM_002582; bases 147–2066) or EGFP were PCR amplified and cloned into the BsrGI and AgeI sites of pLX301 (puromycin resistance; Addgene, 25895) and pLX304 (blasticidin resistance; Addgene, 25890), gifts from D. Root.

Lentiviral vector production and transduction. Lentiviral vectors encoding shRNAs or cDNAs were produced by cotransfection of HEK293T cells with the lentiviral plasmids described above, pCMV_dR8.91 and pCMV_VSV-G, using branched polyethylenimine (Sigma). Supernatants were collected, filtered and frozen in aliquots on day 3–4 after transfection. cDNA-expressing lentiviral vectors were concentrated by centrifugation. Titers, knockdown efficiency and cDNA expression were determined by infecting HEK293T cells for 16–24 h using varying quantities of viral supernatant in the presence of 10 μg/ml protamine sulfate. After 24–36 h, infected cells were selected with 2 μg/ml puromycin or 10 μg/ml blasticidin for 3–5 d, and RNA and/or protein were collected for analysis. For knockdown experiments, cells were transduced with lentivirus and collected 5–6 d after selection in puromycin (2 μg/ml for HEK293 cells, 1 μg/ml for fibroblasts and 0.2 μg/ml for iPSCs). For rescue experiments, control or knockdown cells that were selected for 5–6 d after transduction to express the shRNAs were infected with pLX304-based lentiviral vectors expressing PARN or EGFP and collected 4 d after selection in blasticidin (10 μg/ml for HEK293 cells and 5 μg/ml for fibroblasts).

46. Aubert, G., Baerlocher, G.M., Vulto, I., Poon, S.S. & Lansdorp, P.M. Collapse of telomere homeostasis in hematopoietic cells caused by heterozygous mutations in telomerase genes. *PLoS Genet.* **8**, e1002696 (2012).
47. Park, I.H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
48. Park, I.H., Lerou, P.H., Zhao, R., Huo, H. & Daley, G.Q. Generation of human-induced pluripotent stem cells. *Nat. Protoc.* **3**, 1180–1186 (2008).
49. Warlich, E. *et al.* Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. *Mol. Ther.* **19**, 782–789 (2011).
50. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
51. Magoč, T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
52. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
53. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
54. Patro, R., Duggal, G. & Kingsford, C. Salmon: accurate, versatile and ultrafast quantification from RNA-seq data using lightweight-alignment. *bioRxiv* doi:10.1101/021592 (2015).