HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2015 November 01.

Published in final edited form as:

Nat Genet. 2015 May ; 47(5): 512–517. doi:10.1038/ng.3278.

Exome Sequencing Links Mutations in *PARN* and *RTEL1* with Familial Pulmonary Fibrosis and Telomere Shortening

Bridget D. Stuart[1,2], **Jungmin Choi**[3,4], **Samir Zaidi**[3,4], **Chao Xing**[1], **Brody Holohan**[5], **Rui Chen**[6], **Mihwa Choi**[1], **Pooja Dharwadkar**[6], **Fernando Torres**[6], **Carlos E. Girod**[6], **Jonathan Weissler**[6], **John Fitzgerald**[6], **Corey Kershaw**[6], **Julia Klesney-Tait**[7], **Yolanda Mageto**[8], **Jerry W. Shay**[5], **Weizhen Ji**[3,4], **Kaya Bilguvar**[3,9], **Shrikant Mane**[3,9], **Richard P. Lifton**[3,4,9,10], and **Christine Kim Garcia**[1,6]

[1]Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, Texas, USA

[2]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas, USA

[3]Department of Genetics, Yale School of Medicine, New Haven, Connecticut, USA

[4]Howard Hughes Medical Institute, Yale School of Medicine, New Haven, Connecticut, USA

[5]Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, Texas, USA

[6]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, USA

[7]Department of Internal Medicine, University of Iowa Carver College of Medicine, Iowa City, Iowa, USA

[8]Department of Internal Medicine, University of Vermont College of Medicine, Burlington, Vermont, USA

[9]Yale Center for Genome Analysis, Yale School of Medicine, New Haven, Connecticut, USA

[10]Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA

Abstract

Idiopathic pulmonary fibrosis (IPF) is an age-related disease featuring progressive lung scarring. To elucidate the molecular basis of IPF, we performed exome sequencing of familial pulmonary

Corresponding author: Christine Kim Garcia, MD, PhD, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-8591, Telephone: 214-648-1600, Fax: 214-648-1666, christine.garcia@utsouthwestern.edu.

Accession codes. Whole exome sequencing data has been deposited in dbGAP under accession number phs000774.

Author Contributions

C.K.G, C.X. and R.P.L. conceived, designed and directed the study; J.C., S.Z., C.X., W.J., K.B. and S.M. performed genetic analyses, B.D.S, B.H., R.C., M.C. and J.S. performed and directed experiments, P.D., F.T., C.E.G., J.W., J.F., C.K., J.K.-T., and Y.M. contributed clinical evaluations. All authors approved the final manuscript and contributed critical revisions to its intellectual content.

Competing Financial Interests

The authors declare no competing financial interests.

fibrosis kindreds. Gene burden analysis comparing 78 European cases and 2,816 controls implicated *PARN*, an exoribonuclease with no prior connection to telomere biology or disease, with five novel heterozygous damaging mutations in unrelated cases and none in controls (P-value = 1.3×10^{-8}); mutations were shared by all affected relatives (odds in favor of linkage = 4,096:1). *RTEL1*, an established locus for dyskeratosis congenita, harbored significantly more novel damaging and missense variants at conserved residues in cases than controls ($P = 1.6 \times 10^{-6}$). *PARN* and *RTEL1* mutation carriers had shortened leukocyte telomere lengths and epigenetic inheritance of short telomeres was seen in family members. Together these genes explain ~7% of familial pulmonary fibrosis and strengthen the link between lung fibrosis and telomere dysfunction.

Idiopathic pulmonary fibrosis (IPF) is the prototype of adult-onset interstitial lung disease that preferentially affects males and smokers[1,2]. The disease is progressive, with a life expectancy of 2–3 years after diagnosis. The genetic basis of IPF is incompletely understood. Common variants explain a small fraction of disease risk, including loci near *MUC5B* and *TERT*, the protein component of telomerase[3]. Rare coding mutations in *TERT* are found in ~15% of familial pulmonary fibrosis kindreds and show autosomal dominant transmission with incomplete penetrance[4–6]. Less frequently, rare mutations of large effect are found in genes encoding the human telomerase RNA (*TERC*)[4,5], dyskerin (*DKC1*)[7,8], and surfactant proteins (*SFTPC, SFTPA2*)[9–11]. Some probands of familial pulmonary fibrosis kindreds have short telomere lengths that are unexplained by telomerase mutations[12], suggesting a role for other genes involved in telomere maintenance.

Whole exome sequencing and analysis was performed on genomic DNA samples from 99 probands with familial pulmonary fibrosis of unknown genetic cause (see Methods). Principal component analysis of genotypes from exome data revealed that 79 probands clustered with HapMap subjects of European ancestry while 19 and 1 clustered with subjects of Mexican and African American Ancestry, respectively. Control subjects sequenced and analyzed on the same platforms led to the identification of 2,816 controls of European ancestry (Supplementary Fig. 1). In cases and controls 93.3% and 95.3% of all targeted bases had 8 or more independent reads (Supplementary Tables 1 and 2). With a population disease frequency of familial pulmonary fibrosis of < 1 per 100,000[13], we expected dominant alleles of large effect to be individually very rare in the population. To enrich for variants that are likely to alter the function of encoded proteins, we identified damaging variants (premature termination, frameshift, splice site) and variants that altered positions that were highly conserved across phylogeny. We compared the burden of damaging, and damaging plus missense variants at conserved residues, in protein coding genes; we limited the analysis to novel singletons in cases and controls that were not found in the NHLBI Exome Server (ESP) and 1,000 Genomes databases (Supplementary Table 3). There was a mean of 49.0 novel variants per subject in cases and 50.3 in controls. Since subjects in the NHLBI ESP database included subjects with various cardiopulmonary disease, as a check to ensure that we did not exclude disease-related variants, we also performed these analyses considering all alleles with MAF < 0.1% (Supplementary Table 4). Q-Q plots demonstrated good matching of expected and observed P-values for missense and damaging variants,

while observed P-values for damaging singletons were generally below the expected values owing to the paucity of such variants (Fig. 1).

Two genes surpassed thresholds for genome-wide significance in these analyses ($P < 2.4 \times 10^{-6}$ after accounting for examination of 21,000 genes), while no other gene had P-value < 10^{-4} (Fig. 1, Table 1, Supplementary Tables 3, 4 and 5). Poly(A)-specific Ribonuclease Deadenylation Nuclease *(PARN)*, a 3′ exoribonuclease that has not previously been implicated in disease or telomere maintenance, had 6 damaging variants in probands and 0 in controls ($P = 3.8 \times 10^{-10}$); all of these were confirmed by Sanger sequencing and all were absent in dbSNP, 1,000 Genomes, and NHLBI cohorts (Table 2). Two *PARN* variants were identical and subsequently found to be identical by descent from a recent common ancestor (see below). Retrospectively, after consideration of only 5 independent novel damaging mutations in *PARN* from 78 (not 79) unrelated cases, the result remained highly significant ($P = 1.3 \times 10^{-8}$). The significance of this result is further supported by the absence of damaging mutations in *PARN* among 6,500 subjects (including 4,300 subjects of European ancestry) in the NHLBI controls ($P = 1.7 \times 10^{-9}$ in European cases vs. controls). An additional novel missense variant (Lys421Arg) was found in a proband of European ancestry. *PARN* is among the 20% of genes with the lowest prevalence of rare variants that are likely to disrupt normal function (mutation-intolerant genes)[14], consistent with these variants causing haploinsufficiency.

Two probands not known to be related (F349 and F373) shared the same rare *PARN* variant, altering the canonical splice acceptor of the fourth intron (AG>GG; IVS4 –2a>g). Tracing birth and death records of both kindreds established that these two individuals had a common great grandmother (individual II.2, Fig. 2). Kinship analysis using the Beagle program revealed that the probands share an estimated 6.4% of their genomes. The IVS4 –2a>g variant lies on a segment of identity by descent that is ~18 Mb in length, supporting the relatedness of these two individuals.

As an independent test of the significance of these *PARN* variants, their segregation was compared to the segregation of pulmonary fibrosis in the extended kindreds of each index case. Among 7 relatives with pulmonary fibrosis in whom mutation status was assessed either by direct sequencing or by imputation of obligate carriers, all inherited the novel *PARN* variants identified in their respective probands, an event that was highly unlikely to occur by chance (lod score of 3.6, backwards odds of 4,096:1 in favor of linkage of rare *PARN* variants in affected-only analysis). Due to the rarity of the identified mutant alleles, these lod scores are only modestly changed by increased estimates of mutant allele frequencies. The mutations were also shared by 5 relatives identified as having significant lung disease who did not meet current criteria for a diagnosis of interstitial lung disease (Supplementary Table 6). There were also 9 clinically unaffected subjects who harbored the rare variants found in probands, indicating incomplete penetrance of pulmonary fibrosis.

All five loss-of-function *PARN* variants involve residues within the CAF1 ribonuclease domain, which is conserved through yeast and encodes a critical component of a cytoplasmic deadenylase (Fig. 2g). A lymphoblastic cell line (LCL) derived from the proband with the Gln177* mutation demonstrated greater expression of the wildtype than

the mutant allele (Supplementary Fig. 2a), and PARN protein expression was reduced in six independent LCLs representing three different loss of function mutations (Gln177*, IVS4 and IVS6 splice site) (Supplementary Fig. 2b). There was no apparent decrease in PARN expression in the LCLs derived from subjects heterozygous for the Lys421Arg variant.

The other gene with a significant mutation burden was Regulator of Telomere Elongation Helicase 1 (*RTEL1*), which is known to play a role in telomere maintenance. Mutations in *RTEL1* have recently been shown to cause Hoyeraal-Hreidarsson syndrome, a severe variant of dyskeratosis congenita presenting in childhood and associated with telomere shortening[15–19]. Affected subjects typically have biallelic mutations, however heterozygotes have sometimes been noted to display disease manifestations.

We found five novel heterozygous variants in *RTEL1* (NM_1283009.1) in pulmonary fibrosis probands (two damaging and three missense variants at highly conserved positions), whereas four singletons were observed among 2,816 control subjects (Fig. 3, Table 2, $P = 1.6 \times 10^{-6}$). Similarly, in the NHLBI cohort, there were six singleton variants in *RTEL1* among 4,300 subjects of European ancestry (2 damaging and 4 missense variants at conserved residues; $P = 7.1 \times 10^{-7}$ vs. cases). *RTEL1* ranks among the 2% most mutation-intolerant genes in the genome[14], consistent with phenotypic effects from heterozygous mutations. RTEL1 contains an amino-terminal helicase domain which preserves telomere length during replication by unwinding the repetitive telomere TTAGGG sequences, which are organized in G-quartet secondary structures, and by disassembling the lasso-like T-loops at the end of telomeres[20]. Two rare *RTEL1* variants are predicted to be loss-of-function alleles; the frameshift mutation (Gly201Glufs) and the premature truncation (Gln693*) are predicted to interrupt the highly conserved Rad3-related DNA helicase domain (Fig. 3f). Three missense alleles found in probands affect highly conserved amino acids (Fig. 3g). As seen with *PARN*, all relatives diagnosed with pulmonary fibrosis in whom mutation status could be assessed inherited the same rare *RTEL1* variant found in probands (backward odds of 128:1 in favor of linkage). Five mutation carriers were reported to be free of pulmonary disease, indicative of incomplete penetrance. Pulmonary fibrosis has been previously reported in only one case of severe dyskeratosis congenita due to biallelic *RTEL1* mutations[16], and none of the heterozygous *RTEL1* mutation carriers described herein demonstrate the mucocutaneous features of dyskeratosis congenita (Supplementary Table 6). These mutations thus expand the clinical phenotype associated with *RTEL1* mutations.

Telomere lengths of genomic DNA isolated from circulating leukocytes of *PARN* and *RTEL1* probands were shorter than telomere lengths of unrelated spouses ($P < 0.001$ for both) by terminal restriction fragment length (TRFL) analysis (Fig. 4a, b, c). The *PARN* and *RTEL1* probands also demonstrate a higher percentage of short telomere lengths (TRFL < 3 kb) than normal controls ($P = 0.003$ and $P = 0.001$, respectively). Telomere lengths were also measured using a quantitative PCR assay[6,12]. We used estimated regression coefficients to calculate the age-adjusted telomere lengths of each subject. Six of the seven *PARN* probands and all five *RTEL1* probands had age-adjusted telomere lengths below the 10th percentile (Fig. 4d). The proband with the PARN c.751delA (Arg251Glufs*14) mutation had telomere lengths ~30th percentile. More than half of all mutation carriers studied (24/46) had telomere lengths below the 1st percentile, 65% were below the 10th percentile,

and nearly all were at or below the 50th percentile. The mean age-adjusted telomere lengths of the *PARN* and *RTEL1* mutation carriers were significantly shorter than normal controls (P-value < 0.001 for both; Fig. 4e). They were also similar to telomere length measurements of *TERT* and *TERC* mutation carriers[21]. The mean telomere lengths of related family members who did not inherit the *PARN* or *RTEL1* mutations were shorter than unrelated individuals (i.e. spouses), but longer than *PARN* or *RTEL1* mutation carriers (Fig. 4e). This finding is consistent with epigenetic inheritance of telomere lengths, which has been previously observed in *TERT* kindreds[6].

These studies implicate rare heterozygous loss of function variants in *PARN* and *RTEL1* in pulmonary fibrosis with shortened telomeres. The results show a significantly increased burden of these variants in cases compared to two independent control cohorts. In addition, these mutations have large effects on disease risk as indicated by the high odds ratios for association of these rare variants with disease. While no other genes in this study were significant at genome-wide thresholds, we cannot exclude effects of variants in other genes in small numbers of kindreds. This study demonstrates the utility of exome sequencing of probands to discover novel disease loci of large effect despite complexities such as locus heterogeneity, reduced penetrance, and late onset of disease.

Novel damaging or conserved missense *PARN* and *RTEL1* variants were found in 12% of probands studied. By extrapolation to our complete cohort of familial pulmonary fibrosis kindreds that includes subjects with mutations in known genes, *PARN* and *RTEL1* mutations are found in at least 4 and 3%, respectively. These frequencies are intermediate between those for *TERT* mutations (16%), and *SFTPC* (2%) or *SFTPA2* (1%) mutations.

To our knowledge, this is the first report of a clinical phenotype associated with heterozyous *PARN* mutations. PARN plays a role in translational silencing of maternal gene expression through 5′-cap-dependent deadenylation of mRNAs during oocyte maturation and early development[22,23]. In addition, PARN has been shown to be involved in the maturation of argonaute2-cleaved precursor microRNAs[24] and H/ACA box snoRNAs[25]. The mechanism linking *PARN* mutation to telomere shortening is currently unclear. However, H/ACA snoRNAs are known to associate with four evolutionarily conserved and essential proteins: dyskerin, GAR1, NHP2 and NOP10, and mutations of three of these cause dyskeratosis congenita[26–28]. We speculate that PARN haploinsufficiency may lead to telomere shortening through dysregulation of H/ACA box snoRNAs or the RNA component of human telomerase (*TERC*), which itself contains an H/ACA domain.

Short telomere lengths are disproportionately represented in patients with sporadic IPF[12,29] and are associated with worse survival of IPF patients[21]. The identification of two new genes contributing to this trait strengthens the link between telomere attrition and lung fibrosis and identifies a new gene required for telomere maintenance.

Online Methods

Subjects and specimens

This study was approved by the institutional review board at the University of Texas Southwestern Medical Center. Written informed consent was obtained from all participants. Blood samples were obtained from probands and family members of kindreds with familial interstitial pneumonia[30]. Patient laboratory test results and clinical features are summarized from available clinical records; some records were incomplete or unavailable.

Genomic DNA was isolated from circulating leukocytes using an Autopure (Qiagen) according to the manufacturer's instructions. Genomic DNAs were selected from a cohort of 185 kindreds that were screened to exclude those with rare variants in *TERT* (30), *TERC* (5), *SFTPC* (2) and *SFTPA2* (2). Variants in these four genes were determined by Sanger sequencing either by research investigation[5,6,11] or by CLIA-certified companies (GeneDx, Ambry Genetics, Advanced Diagnostics Laboratory at Children's Medical Center in Dallas TX).

Lymphocytes were isolated from blood drawn in Acid Citrate Dextrose tubes using lymphocyte separation medium (MP Biomedicals). Ebstein-Barr virus (EBV)-transformed lymphoblastoid cell lines (LCLs) were established by incubating ~1×10^6 lymphyocytes, either freshly isolated or previously stored in liquid nitrogen, with EBV shed from a marmoset cell line (B95-8, ATCC) and the mitogen phytohemagglutinin. The cell lines were cultured in RPMI 1640 (HyClone) supplemented with 15% FBS, 2 mM L-glutamine, 1 mM pyruvate, penicillin (100 units/ml) and streptomycin (100 ug/ml), fed twice weekly and expanded to a suspension culture volume of at least 50 mL before storage and use.

Whole-Exome Sequencing and Analysis

Targeted capture was performed using the Nimblegen/Roche Exome V2.0 reagent followed by 74 base paired end sequencing on the Illumina platform. Sequence quality statistics for cases and controls across the whole exome are shown in Supplementary Table 1; sequence coverage statistics for *PARN* and *RTEL1* are shown in Supplementary Table 2. Sequences were aligned to the reference genome and variants were called using SAMtools as previously described[31] and GATK haplocaller. Controls included unaffected parents of offspring with autism and congenital heart disease[31–33] and unaffected subjects sequenced for a study of cardiovascular disease (unpublished). All cases and 2,334 of the 2,816 controls were sequenced by the Yale Center for Genome Analysis; the remaining 482 controls were sequenced at the Cold Spring Harbor Laboratory. All cases and controls (noted as YCGA controls) were sequenced on the same capture and sequencing platform and were analyzed using the same algorithms (SAMtools). The variants identified by this pipeline were used in all subsequent analyses. European ancestry in cases and controls was established by principal component analysis using exome data in combination with data from subjects from the HapMap Consortium[34]. Principal component analysis of genotypes from exome data revealed that 79 probands and 2,816 controls clustered with HapMap subjects of European ancestry. Kinship analysis was performed using BEAGLE[35]; overlapping shared tracks were merged to estimate the maximal length of a shared

haplotype. After it was determined that two of the probands were distantly related to each other (Fig. 2a, discussed in manuscript), only 78 unrelated probands were used in gene burden analyses against ethically matched controls. The frequency of variants were evaluated using exome data deposited within the NHLBI Exome Variant Server (ESP) Database, 1,000 genomes and the Yale exome database (a reference database that is distinct from the Yale controls described above), which includes ~ 20,000 independent alleles. Gene burden analysis was performed comparing the number of novel damaging variants (nonsense, frame-shifts, and splice-site mutations), the number of novel missense variants involving highly conserved residues (found in all or all but one ortholog among 47 species with genome sequencing), and the number of variants with a MAF < 0.1% in each gene for cases and controls. Gene burden analyses using the allele as the unit of analysis are reported in Table 1, Supplementary Tables 3 and 4; results using the individual as the unit of analysis are reported in Supplementary Table 5. More than 400 novel variants called in the Yale dataset had confirmation attempted by PCR amplification and Sanger sequencing, with confirmation of 99.8%[31,32,36]. A second and independent control dataset (noted as NHLBI controls) were evaluated from exome data deposited within the NHLBI Exome Variant Server (ESP) Database. All sequences used in the comparative alignment and the position of amino acid conserved domains were obtained from NCBI (www.ncbi.nlm.nih.gov). The comparison of homologous sequences was based on a ClustalW generated alignment using the default settings (www.ebi.ac.uk/clustalw).

Sanger sequencing of genomic DNA

Direct bidirectional Sanger sequencing of the mutations were performed using an ABI 3700 automated sequencer and BigDye terminator cycle sequencing reagents (Applied Biosystems). PCR conditions and primers used to sequence *PARN* and *RTEL1* are listed in Supplementary Table 7.

Linkage Analysis

As each of the novel variants were not observed among more than 20,000 independent alleles, we assumed that it was not independently introduced but was inherited identical by descent (IBD) through obligate carriers in each pedigree. The affected-only backwards logarithm of odds (LOD) scores were calculated as the logarithm of the likelihood ratio (LR) in favor of linkage considering only the affecteds[37]. Assuming a dominant model in which all affecteds share one allele IBD and assuming that there are no sporadic cases within a multiplex family, the equation simplifies to:

$$LR(total)=\frac{1}{P(IBD|r1)}\times\frac{1}{P(IBD|r2)}\times\frac{1}{P(IBD|r3)}\cdots$$

where r(n) is the relative relationship for each family, n. The denominator is the probability that all affecteds in a kindred share one allele IBD, and it depends on the pedigree structure. For example, the probability that an affected sib-pair shares one allele IBD is 1/2.

RNA Analysis and Immunoblotting

RNA was isolated from pelleted LCLs using Trizol (Ambion). After DNase I treatment, 1st strand cDNA was made using random hexamers and the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Oligonucleotides used to amplify the *PARN* 181-nt product were 5′-TGATGAAAAACGTTCACAGG-3′ and 5′-GGTACATGGCTCTAAATCCA-3′. Suspension culture LCLs were pelleted and lyzed with Cytobuster protein extraction reagent (EMDmillipore) containing protease inhibitor cocktail (Sigma). Protein concentrations were determined using the BCA asssay (Pierce). Equal amounts of cell lysate proteins (10 μg) were added to Laemmli sample buffer with β-mercaptoethanol, run on 10% SDS-PAGE gels and transferred to PVBD membranes. Membranes were incubated in blocking buffer (5% dried non-fat milk in TBST (150 mM NaCl, 10 mM Tris, pH 8, 0.1% Tween 20) for one hour before overnight incubation at 4°C with either rabbit polyclonal antibody to PARN (Abcam 125185; 1:1,000 dilution) or mouse monoclonal antibody to tubulin (Sigma T9026, 1:10,000 dilution). Signals were visualized by chemiluminescence (BioRad) after incubation with horseradish peroxidase-conjugated secondary antibody (Cell Signaling Technology). The intensity of each PARN band was quantified using Image J (http://Imageu.nih.gov/ij).

Telomere Length Measurements

Telomere lengths were independently measured by calculating the telomere restriction fragment length (TRFL) as described[38]. The percentage of short telomeres was calculated by computing the percent signal below 3 kb normalized by molecular weight. Telomere lengths of genomic DNA isolated from circulating leukocytes were measured using the same quantitative PCR assay we previously described[6,12,21] and are reported as logarithm-transformed relative ratio of telomeres to single copy gene (relative T/S). A linear regression model using 201 control participants[12] had previously been used to calculate age-adjusted prediction bands and the estimated regression coefficients were used to calculate the observed minus expected, or age-adjusted, telomere lengths.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the probands and their families for their participation, to Ashley Young for technical excellence, to Sam Nolasco, Kim Stephens and Gentry Wools for their help with blood sample collections, the Yale Center for Genome Analysis for production of exome sequence data, and the Yale Center for Genome Analysis and the McDermott Center Bioinformatics Core for sequence analysis. The authors acknowledge funding support from the following sources: US National Institutes of Health (NIH) National Center for Advancing Translational Sciences grant UL1TR001105; NIH grant U54HG006504 (to the Yale Center for Mendelian Genomics); the Howard Hughes Medical Institute (R.P.L.); NIH grant K12HD068369 (B.D.S.); and NIH grant R01HL093096 (C.K.G.).

References for main text

1. Travis WD, et al. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med. 2013; 188:733–748. [PubMed: 24032382]

2. Raghu G, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011; 183:788–824. [PubMed: 21471066]
3. Fingerlin TE, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet. 2013; 45:613–620. [PubMed: 23583980]
4. Armanios MY, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med. 2007; 356:1317–1326. [PubMed: 17392301]
5. Tsakiri KD, et al. Adult-onset pulmonary fibrosis caused by mutations in telomerase. Proc Natl Acad Sci U S A. 2007; 104:7552–7557. [PubMed: 17460043]
6. Diaz de Leon A, et al. Telomere lengths, pulmonary fibrosis and telomerase (TERT) mutations. PLoS ONE. 2010; 5:e10680. [PubMed: 20502709]
7. Kropski JA, et al. A novel dyskerin (DKC1) mutation is associated with familial interstitial pneumonia. Chest. 2014; 146:e1–7. [PubMed: 24504062]
8. Parry EM, et al. Decreased dyskerin levels as a mechanism of telomere shortening in X-linked dyskeratosis congenita. J Med Genet. 2011; 48:327–333. [PubMed: 21415081]
9. Nogee LM, et al. A mutation in the surfactant protein C gene associated with familial interstitial lung disease. N Engl J Med. 2001; 344:573–579. [PubMed: 11207353]
10. Thomas AQ, et al. Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspecific interstitial pneumonitis in one kindred. Am J Respir Crit Care Med. 2002; 165:1322–1328. [PubMed: 11991887]
11. Wang Y, et al. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. Am J Hum Genet. 2009; 84:52–59. [PubMed: 19100526]
12. Cronkhite JT, et al. Telomere shortening in familial and sporadic pulmonary fibrosis. Am J Respir Crit Care Med. 2008; 178:729–737. [PubMed: 18635888]
13. Marshall RP, Puddicombe A, Cookson WO, Laurent GJ. Adult familial cryptogenic fibrosing alveolitis in the United Kingdom. Thorax. 2000; 55:143–146. [PubMed: 10639533]
14. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013; 9:e1003709. [PubMed: 23990802]
15. Ballew BJ, et al. A recessive founder mutation in regulator of telomere elongation helicase 1, RTEL1, underlies severe immunodeficiency and features of Hoyeraal Hreidarsson syndrome. PLoS Genet. 2013; 9:e1003695. [PubMed: 24009516]
16. Deng Z, et al. Inherited mutations in the helicase RTEL1 cause telomere dysfunction and Hoyeraal-Hreidarsson syndrome. Proc Natl Acad Sci U S A. 2013; 110:E3408–3416. [PubMed: 23959892]
17. Le Guen T, et al. Human RTEL1 deficiency causes Hoyeraal-Hreidarsson syndrome with short telomeres and genome instability. Hum Mol Genet. 2013; 22:3239–3249. [PubMed: 23591994]
18. Walne AJ, Vulliamy T, Kirwan M, Plagnol V, Dokal I. Constitutional mutations in RTEL1 cause severe dyskeratosis congenita. Am J Hum Genet. 2013; 92:448–453. [PubMed: 23453664]
19. Ballew BJ, et al. Germline mutations of regulator of telomere elongation helicase 1, RTEL1, in Dyskeratosis congenita. Hum Genet. 2013; 132:473–480. [PubMed: 23329068]
20. Vannier JB, Pavicic-Kaltenbrunner V, Petalcorin MI, Ding H, Boulton SJ. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. Cell. 2012; 149:795–806. [PubMed: 22579284]
21. Stuart BD, et al. Effect of telomere length on survival in patients with idiopathic pulmonary fibrosis: an observational cohort study with independent validation. The lancet Respiratory medicine. 2014; 2:557–565. [PubMed: 24948432]
22. Korner CG, et al. The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of Xenopus oocytes. The EMBO journal. 1998; 17:5427–5437. [PubMed: 9736620]
23. Dehlin E, Wormington M, Korner CG, Wahle E. Cap-dependent deadenylation of mRNA. The EMBO journal. 2000; 19:1079–1086. [PubMed: 10698948]

24. Yoda M, et al. Poly(A)-specific ribonuclease mediates 3′-end trimming of Argonaute2-cleaved precursor microRNAs. Cell reports. 2013; 5:715–726. [PubMed: 24209750]
25. Berndt H, et al. PARN-dependent trimming. RNA (New York NY. 2012; 18:958–972.
26. Heiss NS, et al. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat Genet. 1998; 19:32–38. [PubMed: 9590285]
27. Vulliamy T, et al. Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. Proc Natl Acad Sci U S A. 2008; 105:8073–8078. [PubMed: 18523010]
28. Walne AJ, et al. Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. Hum Mol Genet. 2007
29. Alder JK, et al. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. Proc Natl Acad Sci U S A. 2008; 105:13051–13056. [PubMed: 18753630]
30. Steele MP, et al. Clinical and pathologic features of familial interstitial pneumonia. Am J Respir Crit Care Med. 2005; 172:1146–1152. [PubMed: 16109978]
31. Zaidi S, et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature. 2013; 498:220–223. [PubMed: 23665959]
32. Sanders SJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012; 485:237–241. [PubMed: 22495306]
33. Iossifov I, et al. De novo gene disruptions in children on the autistic spectrum. Neuron. 2012; 74:285–299. [PubMed: 22542183]
34. International HapMap C et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449:851–861. [PubMed: 17943122]
35. Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. Am J Hum Genet. 2011; 88:173–182. [PubMed: 21310274]
36. Levy D, et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron. 2011; 70:886–897. [PubMed: 21658582]
37. Olson JM. A general conditional-logistic model for affected-relative-pair linkage studies. Am J Hum Genet. 1999; 65:1760–1769. [PubMed: 10577930]
38. Herbert, B-S.; Shay, JW.; Wright, WE. Current Protocols in Cell Biology. Vol. 18.6.1–18.6.20. John Wiley & Sons, Inc; 2003. Analysis of Telomeres and Telomerase.

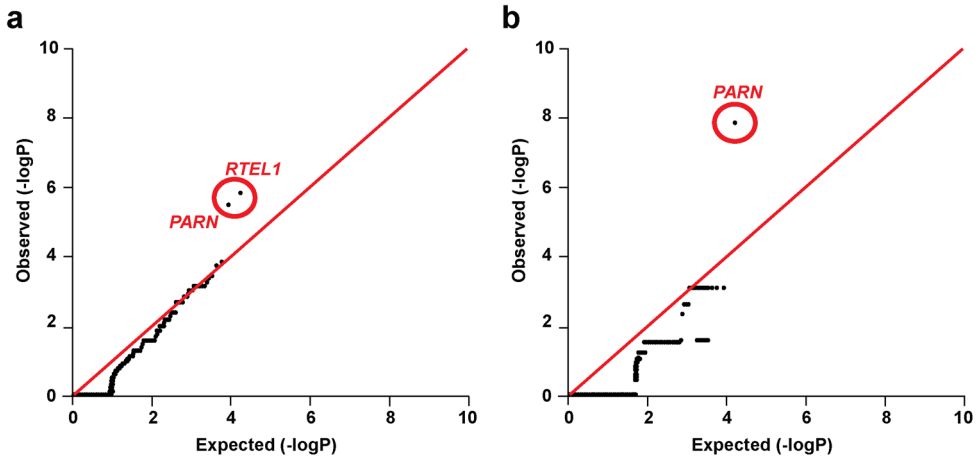

Figure 1. Q-Q Plot of Observed versus Expected P-values Comparing the Burden of Novel Variants in Protein-Coding Genes in Familial Pulmonary Fibrosis Cases and Controls

Novel variants in European 78 pulmonary fibrosis probands and 2,816 controls were identified and their frequencies compared by Fisher's exact test. The distribution of observed P-values for each gene was compared to the distribution of expected P-values. **(a)** Analysis of novel variants that are either damaging or missense at positions that are highly conserved across phylogeny. **(b)** Analysis of novel damaging variants. The distribution of observed P-values generally follows the expected distribution for damaging plus conserved missense variants, while for damaging mutations only, many P-values are lower than expected due to a paucity of variants. The damaging plus missense set shows two genes (*RTEL1* and *PARN*) with P-values at or near genome-wide significance, while the damaging variant set shows one gene, *PARN*, with an observed P-value well outside the expected distribution.

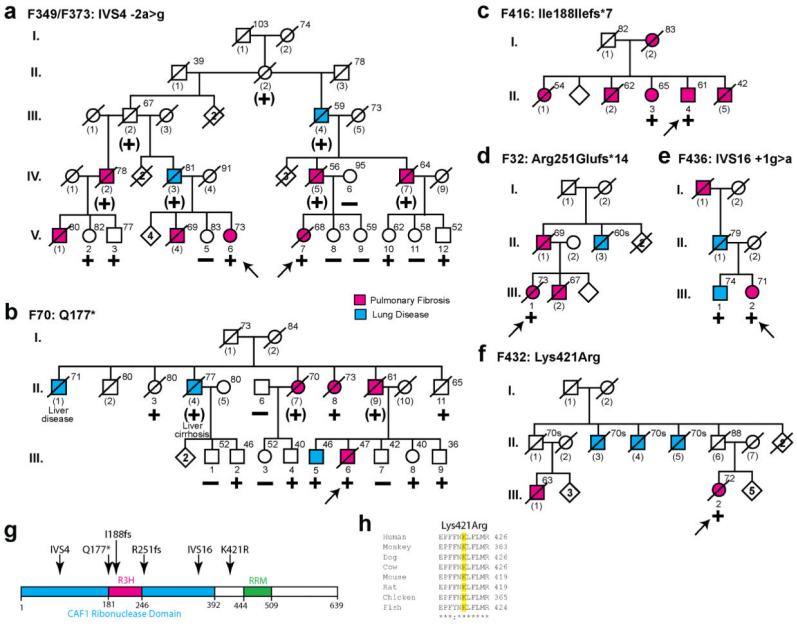

Figure 2. Segregation of Heterozygous *PARN* Mutations in Familial Pulmonary Fibrosis Kindreds and the Location of PARN Alterations in the Different Protein Domains

(a–f) Abridged pedigrees of six kindreds with familial pulmonary fibrosis and *PARN* mutations. The *PARN* cDNA mutations and predicted amino acid changes are listed above each family. Individuals with pulmonary fibrosis or an unclassified lung disease are indicated by red and blue symbols, respectively. Unfilled symbols represent individuals with no self-reported lung disease. Arrows denote the probands. Kindreds F349 and F373 were found to be related through a distant ancestor (II.2). Numbers in parentheses indicate individuals for whom no DNA sample is available. The presence or absence of a mutation is indicated by plus or minus signs, respectively. When the mutation was inferred from location in pedigree, the plus sign is in parentheses. The age at the time of blood draw or death is indicated to the upper right of each symbol. **(g)** Schematic representation of the functional domains of PARN with the position of mutations indicated by the arrows. Conserved protein domains include the CAF1 ribonuclease domain (blue), the R3H domain that binds single stranded nucleic acids (red), and the RNA recognition domain (RRM, green). **(h)** Clustal alignments of homologous PARN protein sequences from *Homo sapiens* (human), *Macaca mulatta* (monkey), *Canis familiaris* (dog), *Bos taurus* (cow), *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Gallus gallus* (chicken), *Danio rerio* (fish).

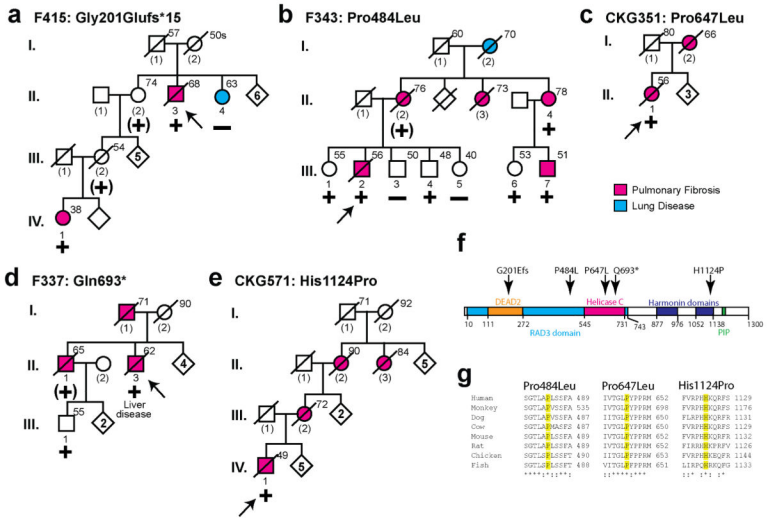

Figure 3. Segregation of Heterozygous *RTEL1* Mutations in Familial Pulmonary Fibrosis Kindreds and the Location of RTEL1 Alterations in the Different Protein Domains

(a–e) Abridged pedigrees of six kindreds with familial pulmonary fibrosis and *RTEL1* mutations. The *RTEL1* cDNA mutations and predicted amino acid changes are listed above each family. Pedigree information is presented in a similar manner as in Figure 1. **(f)** Schematic representation of the functional domains of RTEL1 protein with the position of the mutations indicated by the arrows. Conserved protein domains: DEAD2 domain (orange bar), helicase C domain (red), RAD3-related DNA helicase domain (blue), harmonin N-like domian (purple) and the proliferating cell nuclear antigen (PCNA)-interacting protein (PIP) domain (green). **(g)** Clustal alignments of homologous RTEL1 protein sequences from species listed in Figure 2.

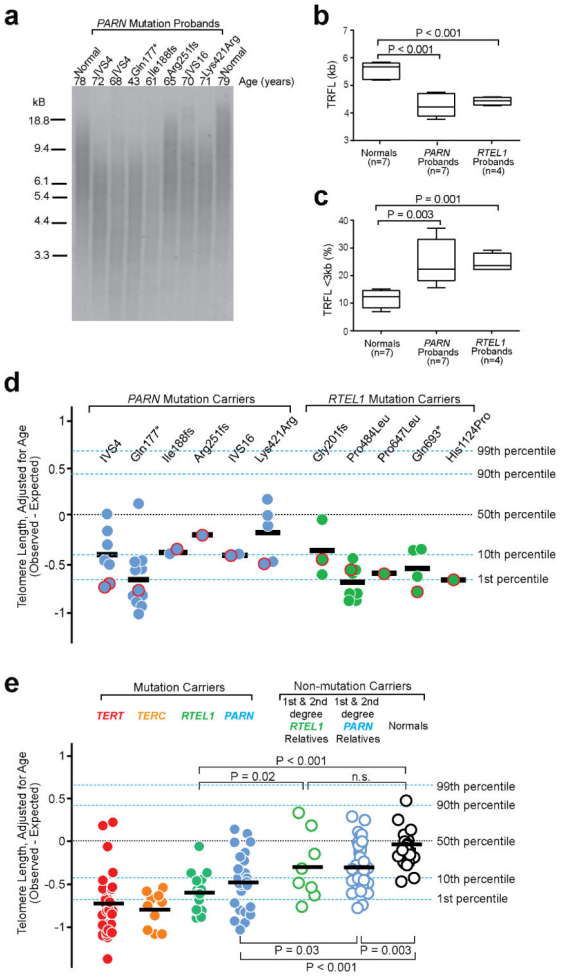

Figure 4. *PARN* and *RTEL1* Mutations Are Associated with Short Leukocyte Telomere Lengths

(a) Telomere lengths determined by Southern blot terminal restriction fragment length (TRFL) analysis of nonrelated normal (spouse) controls and *PARN* probands. The age of each individual is indicated above the blot. Box plots (median, IQR, range) of TRFL **(b)** and percent short TRFL (< 3 kb) **(c)** are indicated for the controls and *PARN* and *RTEL1* probands, who have an average age of 64, 64 and 60 years, respectively. **(d)** Telomere length of genomic DNA isolated from circulating leukocytes from heterozygous *PARN* and *RTEL1* mutation carriers as measured by multiplexed qPCR assay. Telomere lengths are age-adjusted and are expressed as the observed minus the expected length. Each symbol represents a unique individual; the proband for each family is indicated by the red outline. The mean telomere length is indicated by the black bar. The black dotted line shows the median telomere length of an independent control group; approximate age-adjusted prediction bands (percentiles, blue dotted lines) were previously calculated from a linear regression model using the controls. **(e)** Genomic DNA telomere lengths as measured by the multiplexed qPCR assay of individuals heterozygous for a mutation in *TERT*, *TERC*, *RTEL1* and *PARN* are shown as red, orange, green and blue colored circles, respectively. Telomere lengths of 1st and 2nd degree relatives of the *PARN* or *RTEL1* mutation carriers who did not

inherit the variants are shown as open green and blue circles, respectively. Open black circles represent unrelated normal controls.

Table 1

Increased burden of novel variants in *PARN* and *RTEL1* in familial pulmonary fibrosis probands vs. controls of European descent

YCGA controls	Number of Alleles, Damaging plus Missense at Conserved Positions					Number of Alleles, Damaging				
	Controls		Cases		p-value	Controls		Cases		p-value
Gene	Variant	Reference	Variant	Reference		Variant	Reference	Variant	Reference	
PARN	11	5,621	6	150	3.37E–06	0	5,632	5	151	1.33E–08
RTEL1	4	5,598	5	151	1.58E–06	0	5,602	2	154	7.29E–04

NHLBI controls	Number of Alleles, Damaging plus Missense at Conserved Positions					Number of Alleles, Damaging				
	Controls		Cases		p-value	Controls		Cases		p-value
Gene	Variant	Reference	Variant	Reference		Variant	Reference	Variant	Reference	
PARN	8	8,592	6	150	7.76E–08	0	8,600	5	151	1.69E–09
RTEL1	6	8,594	5	151	7.14E–07	2	8,598	2	154	1.85E–03

YCGA controls, variants analyzed from Yale Center for Genome Analysis; NHLBI, variants from NHLBI ESP database

Table 2

Novel *PARN* and *RTEL1* Variants found in Familial Pulmonary Fibrosis Probands by Whole-Exome Sequencing.

Gene	DNA change†	Impact†	dbSNP	1,000 Genome	ESP	Proband Telomere Length (Percentile)
PARN	IVS4 -2a>g‡	Splice	Novel	Novel	Novel	< 1st
PARN	IVS4 -2a>g‡	Splice	Novel	Novel	Novel	< 1st
PARN	c.529 C>T	Gln177*	Novel	Novel	Novel	< 1st
PARN	c.563_564insT	Ile188Ilefs*7	Novel	Novel	Novel	~ 7th
PARN	c.751delA	Arg251Glufs*14	Novel	Novel	Novel	~ 30th
PARN	IVS16 +1g>a	Splice	Novel	Novel	Novel	~ 3rd
PARN	c.1262 A>G	Lys421Arg	Novel	Novel	Novel	< 1st
RTEL1	c.602delG	Gly201Glufs*15	Novel	Novel	Novel	~ 2nd
RTEL1	c.1451 C>T	Pro484Leu	Novel	Novel	Novel	< 1st
RTEL1	c.1940 C>T	Pro647Leu	Novel	Novel	Novel	< 1st
RTEL1	c.2005 C>T	Gln693*	Novel	Novel	Novel	< 1st
RTEL1	c.3371 A>C	His1124Pro	Novel	Novel	Novel	< 1st

†The position of the DNA and protein variants is described using *PARN* NM_2582.3 (variant 1) and *RTEL1* NM_1283009.1 (variant 3).

‡The same mutation was found in two different probands, who are distantly related to each other as shown in Figure 1.

None of the listed variants were found in the dbSNP (version 138), the 1,000 Genome Project (pHase2), or the NHLBI Exome Variant Server Database available from the Exome Sequencing Project (ESP6500SI-V2).