Supplemental Materials for

Utility of Whole Genome Sequencing in Assessing Risk and Clinically-Relevant Outcomes for Pulmonary Fibrosis

Methods

Genome Sequencing and Variant Calling

WGS was performed at the Institute for Genomic Medicine according to standard protocols on Illumina's NovaSeq 6000 platform with 150 bp paired-end reads(1). All samples were processed using the same bioinformatic pipeline for variant calling as previously described(2). Illumina lane-level FASTQ files were aligned to Human Reference Genome (NCBI Build 37) using DRAGEN(3). Sample-level BAM files were archived for bioinformatic telomere length analysis. Picard (http://broadinstitute.github.io/picard) was used to mark duplicates and variant calling was done per the Genome Analysis Toolkit (GATK) Best Practices recommendations (https://gatk.broadinstitute.org/). ClinEff(4) and ATAV(5) (https://github.com/nickzren/atav) was used to further annotate gnomAD frequencies and variant pathogenicity predictors. Samples with >2% contamination by VerifyBamID(6), and those with more than third-degree relatedness were excluded.

Ethnicity Prediction

Using a set of predefined variants previously described(7), we used FlashPCA(8) to perform principal component analysis to define population structure. Principal components 1 and 2 were plotted with overlaying predicted ethnicity (**Figure S1**).

Qualifying Variant Definitions and Annotation

We identified rare telomere-related qualifying variants (QV), restricted to single nucleotide and insertion-deletion (indel) variants from a preset list of telomere-related genes (*TERT, TERC, RTEL1, PARN, DKC1, TINF2, NAF1 and ZCCHC8*). We also identified QVs in non-telomere disease-related genes: *SFTPC*, *SFTPA1, SFTPA2,* and *KIF15*(1, 9). Variants were filtered for rarity: ExAC (Exome Aggregation Consortium release 0.3) or gnomAD (Genome Aggregation Database v2.1) population-specific allele frequency <0.0005 for

African/African-American, Latino/Admixed American, Ashkenazi Jewish, East Asian, South Asian, Finnish, and Non-Finnish European subpopulations. Qualifying deleterious variants comprise protein-truncating variants, including frameshift, stop gained, start lost, or a change in the invariant splice acceptor (GU) or splice donor (AG) sequences, and missense variants with predicted damaging effects from the majority consensus of three *in silico* predictors: Polyphen-2 Humdiv(10) damaging, REVEL(11) > 0.5, and PrimateAI(12) > 0.8. Since the *TERC* gene encodes an RNA template as opposed a translatable transcript, rare *TERC* qualifying variants were annotated as "deleterious" if they had previously been reported to be disease-causing or if the variant has a population-specific allele frequency of <0.0005 and disrupts intramolecular base pairing. All variants were cross-referenced with variant databases in ClinVar(13), Human Gene Mutation Database (HGMD)(14), Online Mendelian Inheritance in Man (OMIM)(15). American College of Medical Genetics and Genomics (ACMG) pathogenicity classification(16) was included in ClinVar.

WGS Telomere Length Estimation

To obtain estimates of telomere length from WGS BAM files (WGS-TL), we used Telseq(17) to quantify telomere repeat containing transcripts while adjusting for GC nucleotide content and total number of sequenced reads.

Polygenic risk score (PRS) calculations

To explore contributions of common variants to WGS-TL, we derived polygenic risk scores from 20 common SNPs previously associated with leukocyte telomere lengths(18) using the following formula:

$$S = \sum_{i=1}^{n} \beta_i X_i$$

In the weighted score, β_i represents the beta coefficient from regression analyses of n conditionally independent signals described previously. X_i represents the number of effect alleles (1 for heterozygous and 2 for homozygous effect alleles, **Table S4**). Since the original coefficients positively correlated with telomere length, we used the additive inverse $(-1 * S)$ to denote risk score for having shortened telomere length. Raw polygenic risk scores ranged from -0.4 to 0.4 and followed a normal distribution with higher scores predicting shorter telomere lengths. Since the raw scores have no intrinsic value they were z-transformed to allow for interpretation of each unit increase in standardized score as one standard deviation increase in raw score.

Generalized Additive Models

We created generalized additive models (GAMs) to test adjusted associations between WGS estimated telomere length (WGS-TL) and common variant polygenic risk scores (PRS) for shortened telomere length. Z-transformed standardized PRS scores were used such that one unit increase corresponded to one standard deviation increase in raw PRS. Each GAM was adjusted for genomic and clinical baseline covariates including age, gender, first two principal components of ethnicity, and presence of rare telomere-related variant (yes/no). We did not make assumptions of linearity of associations between continuous polygenic score and WGS-TL and estimated associations using a nonparametric locally weighted smoothing spline (LOESS). ANOVA for non-parametric effects was used to determine criteria for non-linearity. For independent variables without statistically significant non-linear associations with WGS or qPCR estimated telomere length, a multivariate linear regression was used to estimate effect sizes.

Table S1: Telomere length associated SNPs and effect sizes on telomere length

SNP	Position	Gene	Effect Allele	Effect Allele Freq	β	Std Error	p-value
rs3219104	1:226562621	*PARP1*	C	0.83	0.042	0.006	9.60×10^{-11}
rs10936600	3:169514585	*TERC*	T	0.24	-0.086	0.006	7.18×10^{-51}
rs4691895	4:164048199	*NAF1*	C	0.78	0.058	0.006	1.58×10^{-21}
rs7705526	5:1285974	*TERT*	A	0.33	0.082	0.006	5.34×10^{-45}
rs2853677	5:1287194	*TERT*	A	0.59	-0.064	0.006	3.35×10^{-31}
rs59294613	7:124554267	*POT1*	A	0.29	-0.041	0.006	1.17×10^{-13}
rs9419958	10:105675946	*STN1*	C	0.86	-0.064	0.007	5.05×10^{-19}
rs228595	11:108105593	*ATM*	A	0.42	-0.029	0.005	1.43×10^{-8}
rs2302588	14:73404752	*DCAF4*	C	0.10	0.048	0.008	1.68×10^{-8}
rs7194734	16:82199980	*MPHOSPH6*	T	0.78	-0.037	0.006	6.94×10^{-10}
rs8105767	19:22215441	*ZNF208*	G	0.30	0.039	0.005	5.42×10^{-13}
rs75691080	20:62269750	*RTEL1/ STMN3*	T	0.09	-0.067	0.009	5.99×10^{-14}
rs34978822	20:62291599	*RTEL1*	G	0.02	-0.140	0.023	7.26×10^{-10}
rs73624724	20:62436398	*RTEL1/ ZBTB46*	C	0.13	0.051	0.007	6.33×10^{-12}
rs55749605	3:101232093	*SENP7*	A	0.58	-0.037	0.007	2.45×10^{-8}
rs13137667	4:71774347	*MOB1B*	C	0.96	0.077	0.014	2.43×10^{-8}
rs34991172	6:25480328	*CARMIL1*	G	0.07	-0.061	0.011	6.19×10^{-9}
rs2736176	6:31587561	*PRRC2A*	C	0.31	0.035	0.006	3.53×10^{-10}
rs3785074	16:69406986	*TERF2*	G	0.26	0.035	0.006	4.64×10^{-10}
rs62053580	16:74680074	*RFWD3*	G	0.17	-0.039	0.007	4.08×10^{-8}

Summary statistics including beta and p-values from published GWAS on telomere length (https://doi.org/10.1016/j.ajhg.2020.02.006).

Table S2: Demographic characteristics of patient cohorts

Characteristic	UTSW-IPF (n = 626)	CUMC-IPF (n = 91)	IPFnet (n=232)	p-value
Age (median, IQR)	66 (13)	70 (11)	68 (11)	p < 0.001
Male, n (%)	409 (65%)	64 (70%)	173 (75%)	p = 0.03
Ethnicity, n (%)				
White	516 (82%)	76 (84%)	216 (93%)	p < 0.001
Black	16 (3%)	4 (4%)	3 (1%)	p = 0.3
Hispanic	62 (10%)	0	5 (2%)	p < 0.001
Asian	32 (5%)	11 (12%)	8 (4%)	p = 0.01
Clinical diagnosis*, n (%)				
Idiopathic pulmonary fibrosis	496 (79%)	67 (74%)	232 (100%)	p < 0.001
Unclassifiable lung disease	76 (12%)	9 (10%)	0	p < 0.001
Chronic hypersensitivity pneumonitis	25 (4%)	6 (7%)	0	p < 0.001
Connective tissue disease	12 (2%)	4 (4%)	0	p = 0.008
Idiopathic interstitial pneumonia	15 (2%)	1 (1%)	0	p = 0.03
Other[#]	2 (<1%)	0	0	p = 1
Familial Disease, n (%)	337 (54%)	51 (56%)	ND	

ND: No data

*All non-IPF cases had familial pulmonary fibrosis
[#]Asbestosis (1), sarcoidosis (1)

Table S3. Clinical and Genomic Characteristics by QV carrier gene

Characteristic	All (n = 949)	Telomere-related QV carriers*							Non-telomere QV carriers*			No QV (n = 816)	p-value
		TERT (n = 60)	*TERC* (n = 10)	*RTEL1* (n = 21)	*PARN* (n = 16)	*DKC1* (n = 2)	*TINF2* (n = 1)	*NAF1* (n = 5)	*SFTPC* (n = 8)	*SFTPA* (n = 1)	*KIF15* (n =12)		
Age (median, IQR)	67 (13)	61 (11)	59 (15)	63 (11)	64 (6)	54 (1)		69 (6)	54 (15)		64 (14)	68 (12)	**p < 0.001**
Male, n (%)	646 (68%)	37 (64%)	4 (40%)	11 (52%)	8 (50%)	2 (100%)		5 (100%)	4 (50%)		11 (92%)	563 (69%)	p = 0.05
Ethnicity, n (%)													
White	808 (85%)	54 (90%)	9 (90%)	19 (90%)	14 (88%)	2 (100%)		5 (100%)	7 (88%)		9 (75%)	690 (84%)	NS
Black	23 (3%)	0	0	1 (5%)	0	0		0	0		1 (8%)	21 (3%)	NS
Hispanic	67 (7%)	5 (8%)	0	1 (5%)	1 (6%)	0		0	1 (12%)		2 (17%)	57 (7%)	NS
Asian	51 (5%)	1 (2%)	1 (10%)	0	1 (6%)	0		0	0		0	48 (6%)	NS
IPF Diagnosis, n (%)	795 (84%)	42 (70%)	6 (60%)	13 (62%)	12 (75%)	2 (100%)	1 (100%)	5 (100%)	6 (75%)	1 (100%)	10 (91%)	699 (86%)	**p < 0.01**
Familial Disease†, n (%)	388 (54%)	47 (89%)	9 (100%)	14 (88%)	10 (71%)	1 (50%)	1 (100%)	3 (100%)	7 (100%)	1 (100%)	7 (70%)	291 (48%)	**p < 0.001**
Telomere Length													
WGS-TL (mean±SD)	3.47±0.5	2.9±0.4	3.1±0.6	3.1±0.3	3.2±0.6	3.2±0.3	2.31	3.4±0.6	3.7±0.9	3.2	3.4±0.5	3.5±0.5	**p < 0.001**
qPCR-TL <10th %tile, n (%)	385 (41%)	53 (88%)	8 (80%)	16 (76%)	10 (63%)	2 (100%)	1 (100%)	1 (20%)	3 (38%)	0	3 (27%)	288 (35%)	**p < 0.001**
MUC5b rs35705950 MAF [95% CI]	0.33 [0.31,0.35]	0.24 [0.17,0.33]	0.25 [0.10,0.49]	0.26 [0.14,0.42]	0.22 [0.10,0.40]	0.25 [0.01,0.78]	0.5	0.6 [0.27,0.96]	0.44 [0.21,0.69]	0.5	0.45 [0.25,0.67]	0.34 [0.31,0.36]	**p = 0.02**

Abbreviations: WGS-TL (Whole genome sequencing derived telomere length); IPF (Idiopathic pulmonary fibrosis); MAF (minor allele frequency); NS (not significant)

*****Three individuals carry both a telomere-related QV and a non-telomere QV (*SFTPC/TINF2, SFTPC/PARN, KIF15/TERT*) and included in both groups; For privacy reasons, demographic information for one *SFTPA1/2* QV carrier and one *TINF2* QV carrier not shown

†Number of subjects and proportion for each category restricted to the subjects with known family histories: All (n=717), *TERT* carriers (n=53), *TERC* carriers (n=9), *RTEL1* carriers (n=16), *PARN* carriers (n=14), *DKC1* carriers (n=2), *TINF2* carriers (n=1), *NAF1* carriers (n=3), *SFTPC* carriers (n=7), *SFTPA1/2* carriers (n=1), *KIF15* carriers (n=10), No QV (n=603)

Table S4: Rare qualifying variants identified in IPF/FPF

Gene	Variant	Effect	HGVS coding	HGVS protein	qPCR TL Age-adjusted %tile	WGS TL	Polyphen Humdiv	REVEL	Primate AI	Global AF[&]	Pop AF[$]	ClinVar ClinSig[#]	HGMD DM/DM?^
DKC1	X-153997521-G-A	missense	c.851G>A	p.Arg284Gln	2	2.98	probably	0.691	0.713	NA	NA	NA	Yes
DKC1	X-153999092-A-G	missense	c.974A>G	p.Asp325Gly	1	3.44	probably	0.696	0.785	NA	NA	NA	No
NAF1	4-164048168-G-GT	frameshift	c.1132dupA	p.Thr378fs	34	3.61	NA	NA	NA	0.000159	0.000379	NA	No
NAF1	4-164050411-G-A	stop gained	c.1123C>T	p.Arg375*	3	2.58	NA	NA	NA	NA	NA	NA	No
NAF1	4-164066950-C-G	missense	c.701G>C	p.Arg234Pro	50	3.08	probably	0.545	0.556	NA	NA	NA	No
NAF1	4-164069492-C-A	splice donor	c.634+1G>T	NA	12	3.51	NA	NA	NA	NA	NA	NA	No
NAF1	4-164085483-AGAC-A	disruptive inframe deletion	c.423_425delGTC	p.Ser142del	41	4.05	NA	NA	NA	NA	NA	NA	No
PARN	16-14702141-C-T	stop gained	c.473G>A	p.Trp158*	1	2.37	NA	NA	NA	NA	NA	NA	No
*PARN**	16-14647997-G-A	missense	c.1147C>T	p.Arg383Cys	1	2.50	probably	0.609	0.856	6.37E-05	0.000117	Uncertain significance	Yes
PARN	16-14698005-G-A	stop gained	c.598C>T	p.Gln200*	1	3.06	NA	NA	NA	NA	NA	NA	No
PARN	16-14704526-G-A	stop gained	c.346C>T	p.Gln116*	1	2.64	NA	NA	NA	NA	NA	Pathogenic	Yes
PARN	16-14721046-T-C	splice acceptor	c.63-2A>G	NA	1	3.10	NA	NA	NA	NA	NA	Pathogenic	Yes
PARN	16-14721046-T-C	splice acceptor	c.63-2A>G	NA	2	2.66	NA	NA	NA	NA	NA	Pathogenic	Yes
PARN	16-14647972-A-G	missense	c.1172T>C	p.Phe391Ser	2	2.71	probably	0.704	0.881	NA	NA	NA	No
PARN	16-14540858-CCT-C	frameshift	c.1566_1567delAG	p.Glu524fs	7	2.50	NA	NA	NA	9.56E-05	NA	Pathogenic/ Likely	Yes
PARN	16-14700378-G-A	missense	c.527C>T	p.Pro176Leu	5	3.93	probably	0.577	0.771	NA	NA	Uncertain significance	No
PARN	16-14723964-T-G	missense	c.19A>C	p.Asn7His	9	3.48	probably	0.656	0.697	3.18E-05	NA	Uncertain significance	Yes
PARN	16-14687201-TC-T	frameshift	c.691delG	p.Asp231fs	16	4.05	NA	NA	NA	NA	NA	NA	Yes
PARN	16-14702971-T-TA	frameshift	c.380dupT	p.Glu128fs	24	3.43	NA	NA	NA	NA	NA	Pathogenic	Yes
PARN	16-14698034-CT-C	frameshift	c.568delA	p.Arg190fs	50	3.26	NA	NA	NA	NA	NA	NA	Yes
PARN	16-14678203-C-T	splice donor	c.898+1G>A	NA	50	2.97	NA	NA	NA	NA	NA	NA	Yes
PARN	16-14647921-C-A	splice donor	c.1222+1G>T	NA	51	4.06	NA	NA	NA	NA	NA	Likely pathogenic	No
PARN	16-14649510-C-T	splice donor	c.1135+1G>A	NA	60	4.13	NA	NA	NA	NA	NA	NA	No
RTEL1	20-62321140-C-G	missense	c.2063C>G	p.Ser688Cys	1	2.87	probably	0.77	0.496	3.19E-05	3.19E-05	Uncertain significance	Yes
RTEL1	20-62324564-C-T	stop gained	c.2920C>T	p.Arg974*	1	3.07	NA	NA	NA	3.20E-05	6.25E-05	Pathogenic	Yes

RTEL1	20-62324564-C-T	stop gained	c.2920C>T	p.Arg974*	1	3.30	NA	NA	NA	3.20E-05	6.25E-05	Pathogenic	Yes
RTEL1	20-62324564-C-T	stop gained	c.2920C>T	p.Arg974*	1	2.67	NA	NA	NA	3.20E-05	6.25E-05	Pathogenic	Yes
RTEL1	20-62319093-C-T	missense	c.1451C>T	p.Pro484Leu	1	2.81	probably	0.759	0.688	3.19E-05	6.69E-05	Uncertain significance	Yes
RTEL1	20-62324600-C-T	stop gained	c.2956C>T	p.Arg986*	1	3.09	NA	NA	NA	0.000128	0.000324	Pathogenic/ Likely	Yes
RTEL1	20-62324600-C-T	stop gained	c.2956C>T	p.Arg986*	29	3.42	NA	NA	NA	0.000128	0.000324	Pathogenic/Likely pathogenic	Yes
RTEL1	20-62311300-G-A	splice donor	c.1135+1G>A	NA	1	3.18	NA	NA	NA	NA	NA	Pathogenic	No
RTEL1	20-62319490-A-G	splice acceptor	c.1596-2A>G	NA	1	3.26	NA	NA	NA	NA	NA	NA	No
RTEL1	20-62319737-C-T	missense	c.1720C>T	p.Arg574Trp	1	3.04	probably	0.691	0.470	NA	NA	Uncertain significance	No
RTEL1	20-62321503-TGAC-T	conservative inframe deletion	c.2206_2208delGAC	p.Asp736del	1	2.21	NA	NA	NA	NA	NA	NA	No
RTEL1	20-62320981-C-T	stop gained	c.2005C>T	p.Gln669*	1	2.66	NA	NA	NA	NA	NA	Pathogenic	Yes
RTEL1	20-62326446-A-C	missense	c.3371A>C	p.His1124Pro	1	2.90	probably	0.721	0.512	NA	NA	Pathogenic	Yes
RTEL1	20-62293226-ATTAT-A	frameshift	c.329_332delTTTA	p.Ile110fs	12	2.84	NA	NA	NA	NA	NA	Pathogenic	No
RTEL1	20-62303913-G-A	missense	c.704G>A	p.Arg235His	3	3.27	probably	0.669	0.645	NA	NA	Uncertain significance	No
RTEL1	20-62320484-A-G	missense	c.1877A>G	p.Lys626Arg	5	3.47	probably	0.748	0.612	NA	NA	Uncertain significance	No
RTEL1	20-62320916-C-T	missense	c.1940C>T	p.Pro647Leu	2	2.85	probably	0.914	0.659	NA	NA	Uncertain significance	Yes
RTEL1	20-62297418-CG-C	frameshift	c.602delG	p.Gly201fs	15	3.38	NA	NA	NA	NA	NA	Pathogenic	Yes
RTEL1	20-62326259-T-C	missense	c.3275T>C	p.Leu1092Pro	18	3.51	probably	0.56	0.403	NA	NA	NA	No
RTEL1	20-62298907-G-T	splice donor	c.699+1G>T	NA	8	3.55	NA	NA	NA	NA	NA	NA	No
RTEL1	20-62320936-C-A	missense	c.1960C>A	p.Pro654Thr	26	3.35	probably	0.799	0.514	NA	NA	NA	No
SFTPA2	10-81317020-C-A	missense	c.692G>T	p.Gly231Val	82	3.28	probably	0.55	0.554	NA	NA	Pathogenic	Yes
SFTPC	8-22020127-G-A	missense	c.83G>A	p.Cys28Tyr	2	2.66	probably	0.904	0.724	NA	NA	NA	No
SFTPC	8-22020156-A-ATCGTGGTGGTGGTGGTGTCCTCATCG	conservative inframe insertion	c.121_147dupGTGGTGGTGGTCCTCATCGTCGTGGTG	p.Val41_Val49dup	50	3.95	NA	NA	NA	NA	NA	NA	No
SFTPC	8-22020220-A-G	missense	c.176A>G	p.His59Arg	1	2.31	probably	0.923	0.720	6.37E-05	0.000405	Likely benign	Yes
SFTPC	8-22020609-T-C	missense	c.218T>C	p.Ile73Thr	50	3.47	probably	0.731	0.564	NA	NA	Pathogenic	Yes
SFTPC	8-22020609-T-C	missense	c.218T>C	p.Ile73Thr	98	4.94	probably	0.731	0.564	NA	NA	Pathogenic	Yes
SFTPC	8-22020609-T-C	missense	c.218T>C	p.Ile73Thr	50	3.96	probably	0.731	0.564	NA	NA	Pathogenic	Yes

SFTPC	8-22020609-T-C	missense	c.218T>C	p.Ile73Thr	8	3.49	probably	0.731	0.564	NA	NA	Pathogenic	Yes
SFTPC	8-22020695-G-A	missense	c.304G>A	p.Val102Met	50	4.45	probably	0.816	0.481	NA	NA	NA	Yes
TERT	5-1253913-G-A	missense	c.3329C>T	p.Thr1110Met	1	3.51	probably	0.509	0.382	3.19E-05	0.00022	Uncertain significance	Yes
TERT	5-1254503-AGTGGCAC-A	frameshift	c.3268_3274delGTGCCAC	p.Val1090fs	1	2.69	NA	NA	NA	NA	NA	NA	No
TERT	5-1254576-C-T	missense	c.3202G>A	p.Glu1068Lys	1	2.87	probably	0.546	0.518	NA	NA	NA	Yes
TERT	5-1254591-C-T	missense	c.3187G>A	p.Gly1063Ser	1	2.50	probably	0.625	0.410	NA	NA	Uncertain significance	Yes
TERT	5-1258753-AC-A	frameshift	c.2991delG	p.Cys998fs	1	3.11	NA	NA	NA	NA	NA	Pathogenic	Yes
TERT	5-1260624-G-A	missense	c.2935C>T	p.Arg979Trp	21	3.71	probably	0.602	0.357	NA	NA	Uncertain significance	Yes
TERT	5-1260647-C-T	missense	c.2912G>A	p.Arg971His	1	3.18	probably	0.516	0.426	NA	NA	Uncertain significance	Yes
TERT	5-1260690-T-G	missense	c.2869A>C	p.Ser957Arg	1	2.82	probably	0.749	0.577	NA	NA	Uncertain significance	Yes
TERT	5-1264550-G-A	missense	c.2812C>T	p.Arg938Trp	1	3.07	probably	0.614	0.515	NA	NA	Likely pathogenic	Yes
TERT	5-1264587-G-T	missense	c.2775C>A	p.His925Gln	50	3.43	probably	0.617	0.635	NA	NA	Uncertain significance	Yes
TERT	5-1266586-A-T	missense	c.2647T>A	p.Phe883Ile	1	3.40	probably	0.888	0.647	NA	NA	NA	Yes
TERT	5-1266612-G-C	missense	c.2621C>G	p.Thr874Arg	1	2.16	probably	0.879	0.523	NA	NA	NA	Yes
TERT	5-1266634-C-T	missense	c.2417G>A	p.Gly806Asp	6	3.05	probably	0.654	0.333	NA	NA	Uncertain significance	Yes
TERT	5-1266639-C-T	missense	c.2594G>A	p.Arg865His	3	3.20	probably	0.939	0.506	3.19E-05	0.000149	Conflicting interpretation	Yes
TERT	5-1266639-C-T	missense	c.2594G>A	p.Arg865His	1	3.34	probably	0.939	0.506	3.19E-05	0.000149	Conflicting interpretation	Yes
TERT	5-1266652-T-A	splice acceptor	c.2401-2A>T	NA	1	2.32	NA	NA	NA	NA	NA	Pathogenic	Yes
TERT	5-1268636-C-T	missense +splice region	c.2581G>A	p.Gly861Arg	1	3.34	probably	0.801	0.440	NA	NA	Uncertain significance	Yes
TERT	5-1268696-G-A	missense	c.2339C>T	p.Ala780Val	3	2.79	probably	0.912	0.492	NA	NA	NA	Yes
TERT	5-1268744-A-G	missense	c.2291T>C	p.Leu764Pro	1	3.25	probably	0.855	0.541	NA	NA	NA	No
TERT	5-1279431-G-A	missense	c.2105C>T	p.Pro702Leu	1	3.16	probably	0.619	0.345	NA	NA	Uncertain significance	Yes
TERT	5-1271269-GCG-ACA	missense	c.2431_2433delCGCinsTGT	p.Arg811Cys	1	2.61	probably	0.587	0.242	3.18E-05	3.18E-05	NA	No
TERT	5-1271304-C-T	missense	c.2398G>A	p.Glu800Lys	16	3.63	probably	0.556	0.467	NA	NA	NA	No
TERT	5-1272305-C-T	missense	c.2377G>A	p.Glu793Lys	1	3.50	probably	0.79	0.593	NA	NA	Uncertain significance	No
TERT	5-1278801-GA-G	frameshift	c.2240delT	p.Val747fs	28	2.77	NA	NA	NA	NA	NA	Pathogenic	Yes
TERT	5-1278817-C-T	missense	c.2225G>A	p.Arg742His	1	2.98	probably	0.706	0.468	NA	NA	Uncertain significance	Yes

TERT	5-1279426-G-A	missense	c.2110C>T	p.Pro704Ser	1	2.98	probably	0.619	0.318	3.19E-05	3.19E-05	Conflicting interpretation	Yes
TERT	5-1279426-G-A	missense	c.2110C>T	p.Pro704Ser	1	2.96	probably	0.619	0.318	3.19E-05	3.19E-05	Conflicting interpretation	Yes
TERT	5-1279426-G-A	missense	c.2110C>T	p.Pro704Ser	1	2.83	probably	0.619	0.318	3.19E-05	3.19E-05	Conflicting interpretation	Yes
TERT	5-1279455-A-T	missense	c.2081T>A	p.Val694Glu	1	3.05	probably	0.681	0.577	NA	NA	Uncertain significance	Yes
TERT	5-1279456-C-T	missense	c.2080G>A	p.Val694Met	3	3.45	probably	0.641	0.596	NA	NA	Uncertain significance	Yes
TERT	5-1279503-G-T	missense	c.2033C>A	p.Ala678Asp	1	3.08	probably	0.65	0.608	NA	NA	NA	Yes
TERT	5-1279525-G-A	missense	c.2011C>T	p.Arg671Trp	1	2.75	probably	0.58	0.418	3.19E-05	3.19E-05	Conflicting interpretation	Yes
TERT	5-1280328-G-A	missense	c.1895C>T	p.Pro632Leu	1	2.22	probably	0.835	0.591	NA	NA	Uncertain significance	Yes
TERT	5-1280328-G-A	missense	c.1895C>T	p.Pro632Leu	1	2.58	probably	0.835	0.591	NA	NA	Uncertain significance	Yes
TERT	5-1280331-C-T	missense	c.1892G>A	p.Arg631Gln	1	2.14	probably	0.912	0.447	NA	NA	Pathogenic	Yes
TERT	5-1280331-C-T	missense	c.1892G>A	p.Arg631Gln	1	2.33	probably	0.912	0.447	NA	NA	Pathogenic	Yes
TERT	5-1280332-G-A	missense	c.1891C>T	p.Arg631Trp	3	3.12	probably	0.853	0.565	NA	NA	Pathogenic/ Likely	Yes
TERT	5-1282603-C-A	missense	c.1710G>T	p.Lys570Asn	1	2.79	probably	0.813	0.563	NA	NA	Pathogenic	Yes
TERT	5-1282710-G-A	missense	c.1603C>T	p.Arg535Cys	1	2.88	probably	0.567	0.411	NA	NA	Uncertain significance	No
TERT	5-1282712-T-C	missense	c.1601A>G	p.His534Arg	1	2.29	probably	0.616	0.563	NA	NA	NA	No
TERT	5-1293545-G-A	missense	c.1456C>T	p.Arg486Cys	1	3.10	probably	0.714	0.575	NA	NA	Uncertain significance	Yes
TERT	5-1293545-G-A	missense	c.1456C>T	p.Arg486Cys	5	2.80	probably	0.714	0.575	NA	NA	Uncertain significance	Yes
TERT	5-1293604-C-G	missense	c.1397G>C	p.Arg466Pro	1	2.18	probably	0.824	0.788	NA	NA	NA	Yes
TERT	5-1293951-CAG-C	frameshift	c.1048_1049delCT	p.Leu350fs	1	2.99	NA	NA	NA	NA	NA	Pathogenic	No
TERT	5-1293996-TGAG-T	disruptive inframe deletion	c.1002_1004delCTC	p.Ser335del	1	3.43	NA	NA	NA	NA	NA	Uncertain significance	Yes
TERT	5-1294555-A-T	missense	c.446T>A	p.Leu149Gln	1	3.21	probably	0.739	0.913	NA	NA	NA	No
TERT	5-1294571-C-T	missense	c.430G>A	p.Val144Met	1	2.76	probably	0.667	0.907	NA	NA	Conflicting interpretation	Yes
TERT	5-1294571-C-T	missense	c.430G>A	p.Val144Met	1	2.39	probably	0.667	0.907	NA	NA	Conflicting interpretation	Yes
TERT	5-1294585-A-C	missense	c.416T>G	p.Leu139Arg	1	3.15	probably	0.639	0.872	NA	NA	Uncertain significance	Yes
TERT	5-1294624-G-T	missense	c.377C>A	p.Thr126Lys	3	3.23	probably	0.593	0.931	NA	NA	NA	Yes
TERT	5-1294664-C-CG	frameshift	c.336dupC	p.Glu113fs	3	3.30	NA	NA	NA	NA	NA	Pathogenic	Yes
TERT	5-1294672-C-G	missense	c.329G>C	p.Gly110Ala	14	3.02	probably	0.276	0.803	NA	NA	Uncertain significance	Yes
TERT	5-1294693-AG-CC	missense	c.307_308delCTinsGG	p.Leu103Gly	1	3.23	probably	0.713	0.900	NA	NA	Uncertain significance	No

TERT	5-1294708-G-T	missense	c.293C>A	p.Ala98Asp	1	2.41	probably	0.868	0.947	NA	NA	NA	Yes
TERT	5-1294771-A-G	missense	c.230T>C	p.Leu77Pro	41	3.42	possibly	0.735	0.964	NA	NA	Conflicting interpretation	Yes
TERT	5-1294773-G-T	stop gained	c.228C>A	p.Cys76*	14	3.25	NA	NA	NA	NA	NA	NA	No
TERT	5-1294912-G-T	missense	c.193C>A	p.Pro65Thr	4	3.32	probably	0.401	0.855	NA	NA	Uncertain significance	Yes
TERT	5-1294978-C-T	missense	c.127G>A	p.Asp43Asn	6	2.85	probably	0.576	0.926	NA	NA	Uncertain significance	No
TERT	5-1295008-G-A	missense	c.97C>T	p.Pro33Ser	1	2.47	probably	0.612	0.835	NA	NA	Pathogenic	Yes
TERT	5-1295022-A-C	missense	c.83T>G	p.Val28Gly	1	2.57	probably	0.555	0.921	NA	NA	Uncertain significance	No
TINF2	14-24709848-T-G	missense	c.733A>C	p.Lys245Gln	1	2.31	probably	0.698	0.504	NA	NA	Pathogenic	Yes
TERC	3-169482812-T-C	non coding transcript exon	n.24A>G	NA	2	3.18	NA	NA	NA	3.19E-05	0.000145	Uncertain significance	Yes
TERC	3-169482733-G-A	non coding transcript exon	n.103C>T	NA	1	2.70	NA	NA	NA	NA	NA	Conflicting interpretation	Yes
TERC	3-169482783-G-T	non coding transcript exon	n.53C>A	NA	1	2.64	NA	NA	NA	NA	NA	Uncertain significance	Yes
TERC	3-169482667-C-G	non coding transcript exon	n.169G>C	NA	1	2.97	NA	NA	NA	NA	NA	Uncertain significance	Yes
TERC	3-169482415-C-A	non coding transcript exon	n.421G>T	NA	16	2.81	NA	NA	NA	NA	NA	Uncertain significance	No
TERC	3-169482615-G-C	non coding transcript exon	n.221C>G	NA	1	3.05	NA	NA	NA	NA	NA	Uncertain significance	Yes
TERC	3-169482528-G-GGCTGACA	non coding transcript exon	n.301_307dupTGTCAGC	NA	1	2.19	NA	NA	NA	NA	NA	NA	No
TERC	3-169482606-G-A	non coding transcript exon	n.230C>T	NA	27	4.40	NA	NA	NA	NA	NA	Uncertain significance	No
TERC	3-169482416-C-G	non coding transcript exon	n.420G>C	NA	1	2.86	NA	NA	NA	NA	NA	NA	Yes
TERC	3-169482647-A-G	non coding transcript exon	n.189T>C	NA	8	3.66	NA	NA	NA	NA	NA	Uncertain significance	No
KIF15	3-44816777-C-T	stop gained	c.94C>T	p.Arg32*	18	2.96	NA	NA	NA	6.40E-05	9.62E-05	NA	No
KIF15	3-44816777-C-T	stop gained	c.94C>T	p.Arg32*	50	3.86	NA	NA	NA	6.40E-05	9.62E-05	NA	No
KIF15	3-44816778-G-A	missense	c.95G>A	p.Arg32Gln	1	3.05	probably	0.807	0.661	9.59E-05	0.000163	NA	No
KIF15	3-44816847-C-T	missense	c.164C>T	p.Ser55Phe	20	3.25	probably	0.518	0.533	NA	NA	NA	No
KIF15	3-44819604-TAGG-T	frameshift	c.247-2_247delAGG	p.Glu83fs	50	3.93	NA	NA	NA	3.18E-05	NA	NA	No

KIF15	3-44826336-G-A	splice acceptor	c.362-1G>A	NA	78	3.88	NA	NA	NA	3.41E-05	0.000106	NA	No
KIF15	3-44826384-C-T	missense	c.409C>T	p.Pro137Ser	13	3.41	probably	0.846	0.808	3.22E-05	3.28E-05	NA	No
KIF15	3-44827963-ACT-A	frameshift	c.539_540delCT	p.Ser180fs	4	2.89	NA	NA	NA	NA	NA	NA	No
KIF15	3-44835742-G-A	missense	c.673G>A	p.Ala225Thr	10	2.85	probably	0.804	0.714	9.56E-05	7.04E-05	NA	No
KIF15	3-44844375-CTG-C	frameshift	c.1580_1581delTG	p.Leu527fs	50	4.29	NA	NA	NA	NA	NA	NA	No
KIF15	3-44872495-TTC-T	frameshift	c.3158_3159delCT	p.Ser1053fs	50	3.45	NA	NA	NA	NA	8.80E-06	NA	No
KIF15	3-44882564-A-G	splice acceptor	c.3421-2A>G	NA	50	3.78	NA	NA	NA	NA	0.000131	NA	No

Abbreviations: TL, telomere length; HGVS, Human Genome Variation Society; WGS, whole genome sequencing; HGMD, Human Gene Mutation Database

*Variant found in homozygous state; all other variants are heterozygous
[&]gnomAD v2.1.1 genomes global allele frequency; NA – not found in database
[$]Max allele frequency in gnomAD v2.1.1 exomes and ExAC subpopulations: Ashkenazi Jewish (gnomAD only), European (non-Finnish), African/African-American, Latino/Admixed American, East Asian, European (Finnish), South Asian; NA – not found in database
[#]ACMG classification of variant in ClinVar; NA – not found in database
^HGMD classification; DM, disease-causing; DM?, probable/possible pathogenic mutation

Three subjects had both a telomere-related QV and a non-telomere-related QV (*TINF2/SFTPC*, *PARN/SFTPC*, *TERT/KIF15*); Both carriers of *DKC1* were male

Human reference genome hg19 used for genomic coordinates

Table S5: Baseline characteristics of subgroup with clinical outcomes

	CUMC (n=77)	UTSW (n=397)	P-value
Age, mean (SD)	68.8 (9.7)	66.6 (9.9)	0.07
Male, n (%)	54 (70)	272 (69)	0.88
Non-Hispanic White, n (%)	64 (83)	332 (84)	1
Ever smoker, n (%)	43 (57)	247 (64)	0.3
Familial, n (%)	37 (48)	173 (44)	0.55
Disease Severity at Enrollment			
FVC% predicted, mean (SD)	70.6 (19.9)	65.8 (20.4)	0.06
DLCO% predicted, mean (SD)	41.8 (15.2)	41.6 (19.3)	0.95
GAP score, median (IQR)	4.0 (3.0, 5.0)	4.0 (3.0, 5.0)	0.83
GAP index, median (IQR)	2.0 (1.0, 2.0)	2.0 (1.0, 2.0)	0.7
Clinical Outcomes, n (%)			
Death	6 (8)	197 (50)	**<0.01**
Transplant	9 (12)	97 (24)	**0.02**
FVC >10% decline	28 (36)	221 (56)	**<0.01**
DLCO >15% decline	31 (40)	213 (54)	**0.04**
Follow-up Years, median (IQR)	1.5 (0.7, 2.2)	1.8 (0.8, 3.6)	**<0.01**

Table S6: Baseline characteristics of subgroup with spirometric data

	CUMC (n=41)	UTSW (n=227)	P-value
Age, mean (SD)	68.9 (7.5)	65.5 (10.1)	**0.04**
Male, n (%)	33 (80)	149 (66)	0.09
Non-Hispanic White, n (%)	33 (80)	190 (84)	0.78
Ever smoker, n (%)	19 (46)	132 (60)	0.16
Familial, n (%)	20 (49)	111 (49)	1.0
Disease Severity at Enrollment			
FVC% predicted, mean (SD)	73.4 (20.0)	71.0 (20.2)	0.49
DLCO% predicted, mean (SD)	43.7 (15.5)	48.3 (19.5)	0.15
GAP score, median (IQR)	4 (3-5)	4 (2-5)	0.06
GAP index, median (IQR)	2 (1-2)	2 (1-2)	0.11
Clinical Outcomes, n (%)			
Death	1 (2)	87 (38)	**<0.01**
Transplant	6 (15)	55 (24)	0.25
FVC >10% decline	25 (61)	172 (76)	0.07
DLCO >15% decline	23 (56)	172 (76)	**0.02**
Follow-up Years, median (IQR)	1.8 (1.4-2.6)	3.0 (1.6-4.4)	**<0.01**

Figure S1. Predicted ethnicity estimated from principal components of 12k ancestrally informative SNPs. Classification based on probabilities of > 0.75 for each predicted ethnicity. Admixed samples did not reach > 0.75 probability for any specific ethnicity.

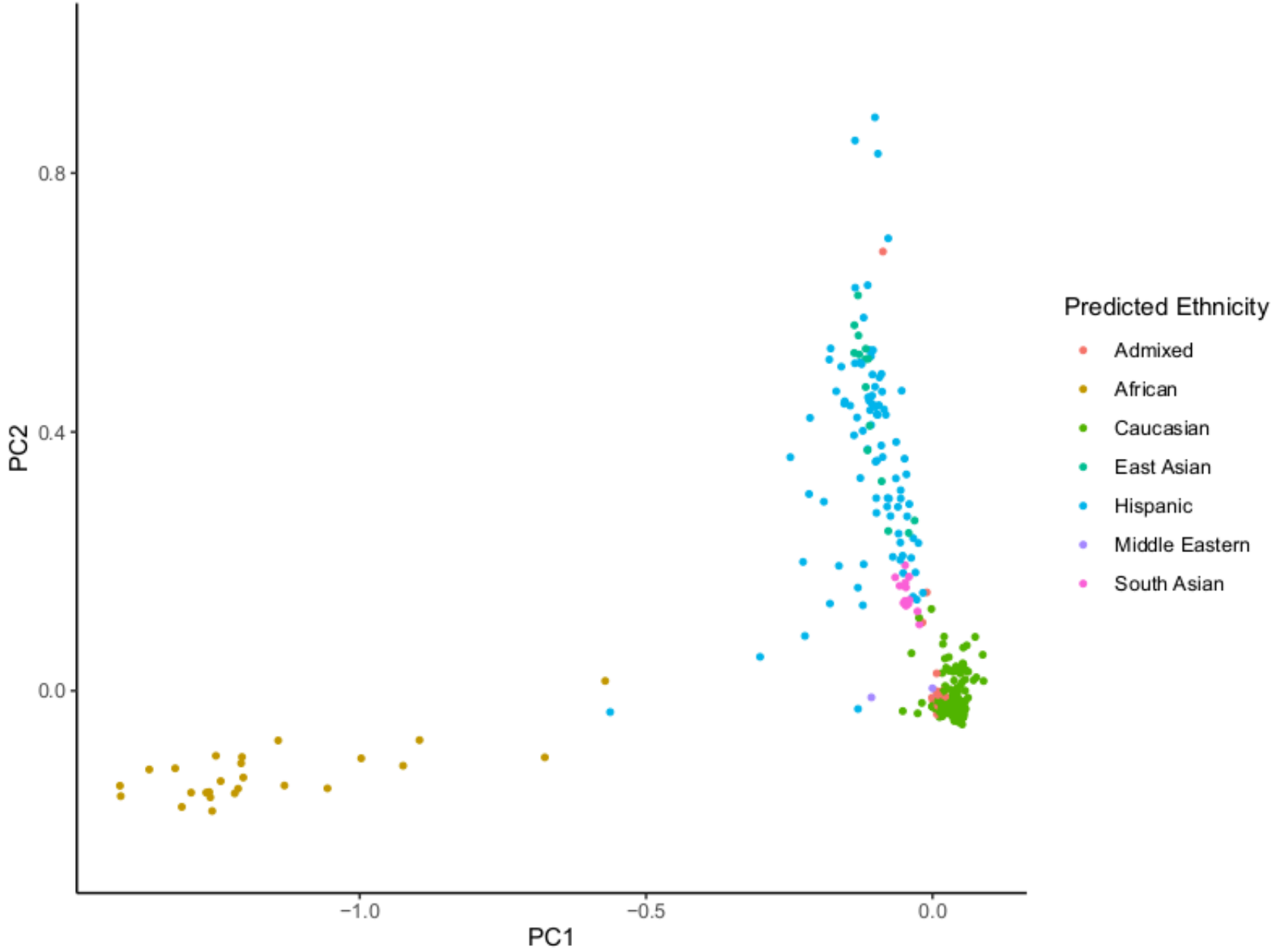


Figure S2. Telomere length of qualifying variant carriers by variant type. qPCR in units of ln(t/s). Protein truncating variants includes stop gained, start lost, frameshift, and splice donor/acceptor site variation. Compared to non-carriers ***p<0.001, **p<0.01, *p<0.05 by Dunnett's test for pairwise comparison with multiple comparison correction.

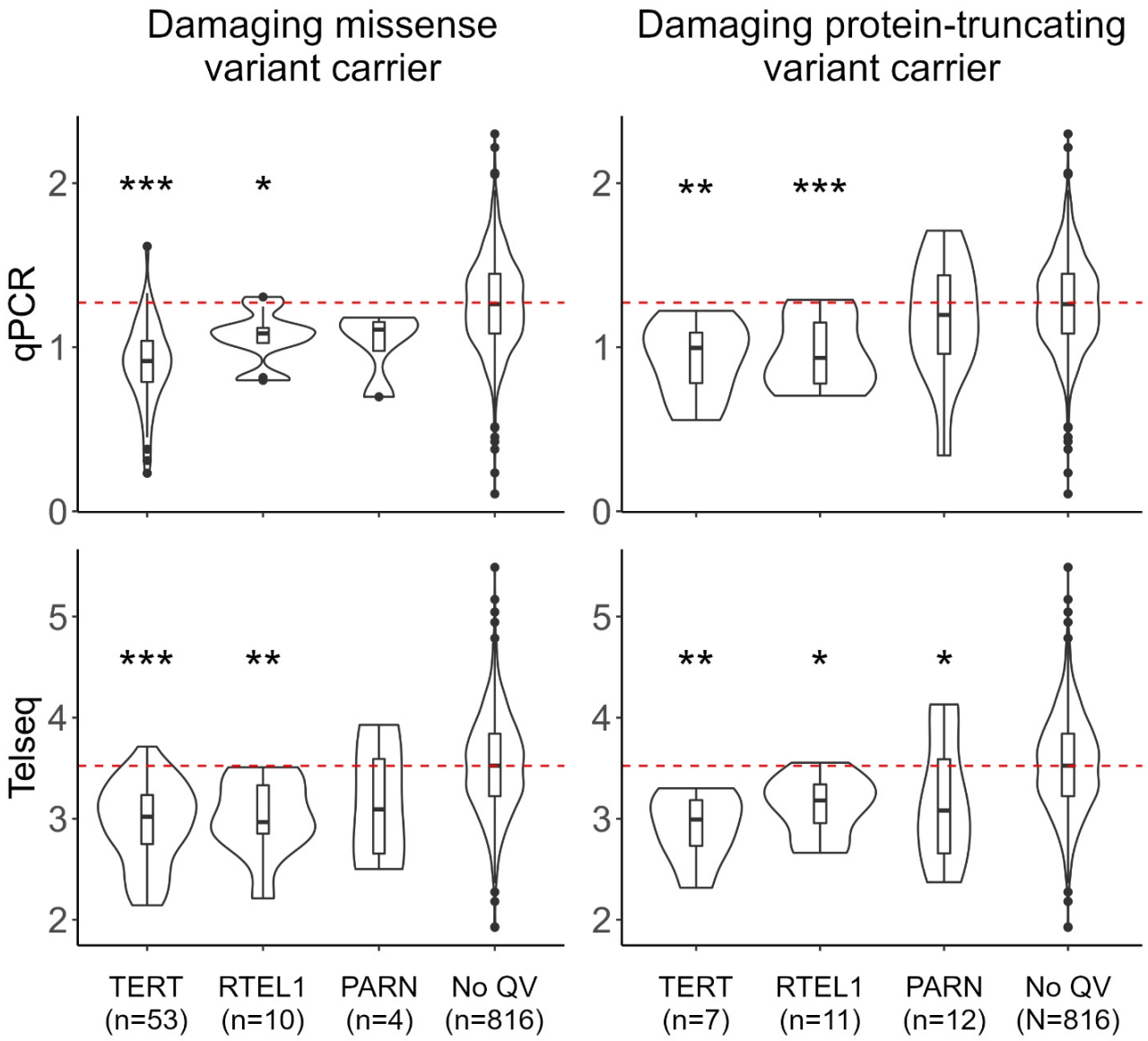


Figure S3. Telomere length of telomere-related qualifying variant carriers by ACMG classification. Genes included are *TERT, TERC, RTEL1, PARN, DKC1, TINF2, NAF1*. Compared to non-carriers ***p<0.001 by Dunnett's test for pairwise comparison with multiple comparison correction.

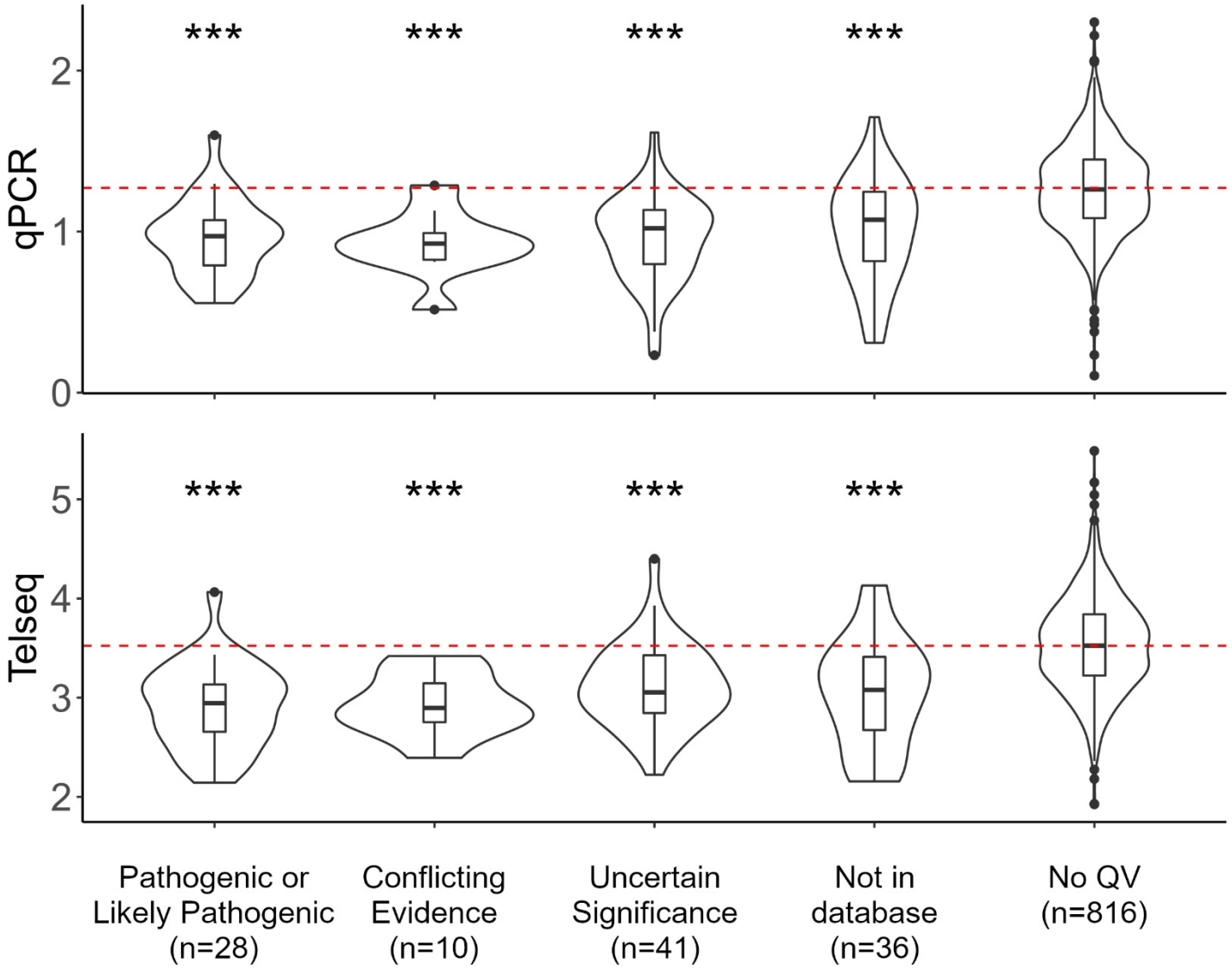


Figure S4. STROBE diagram of genetic, survival, and progression analyses.

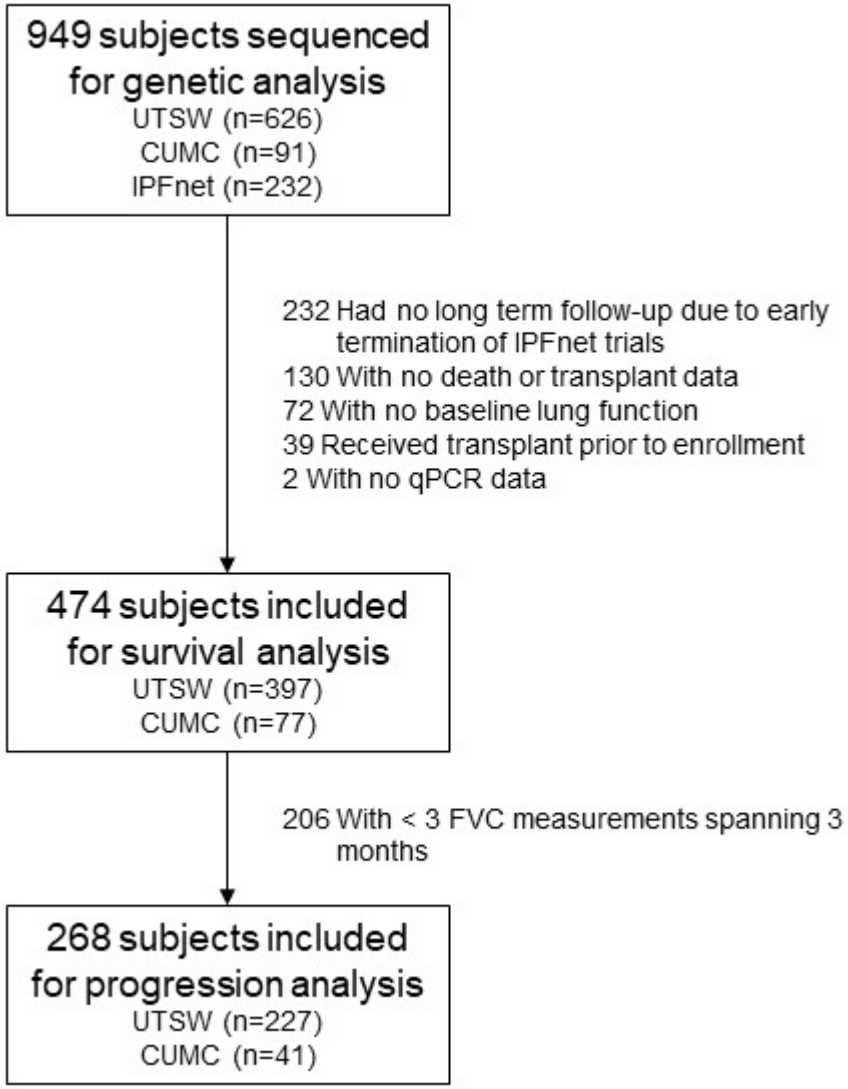
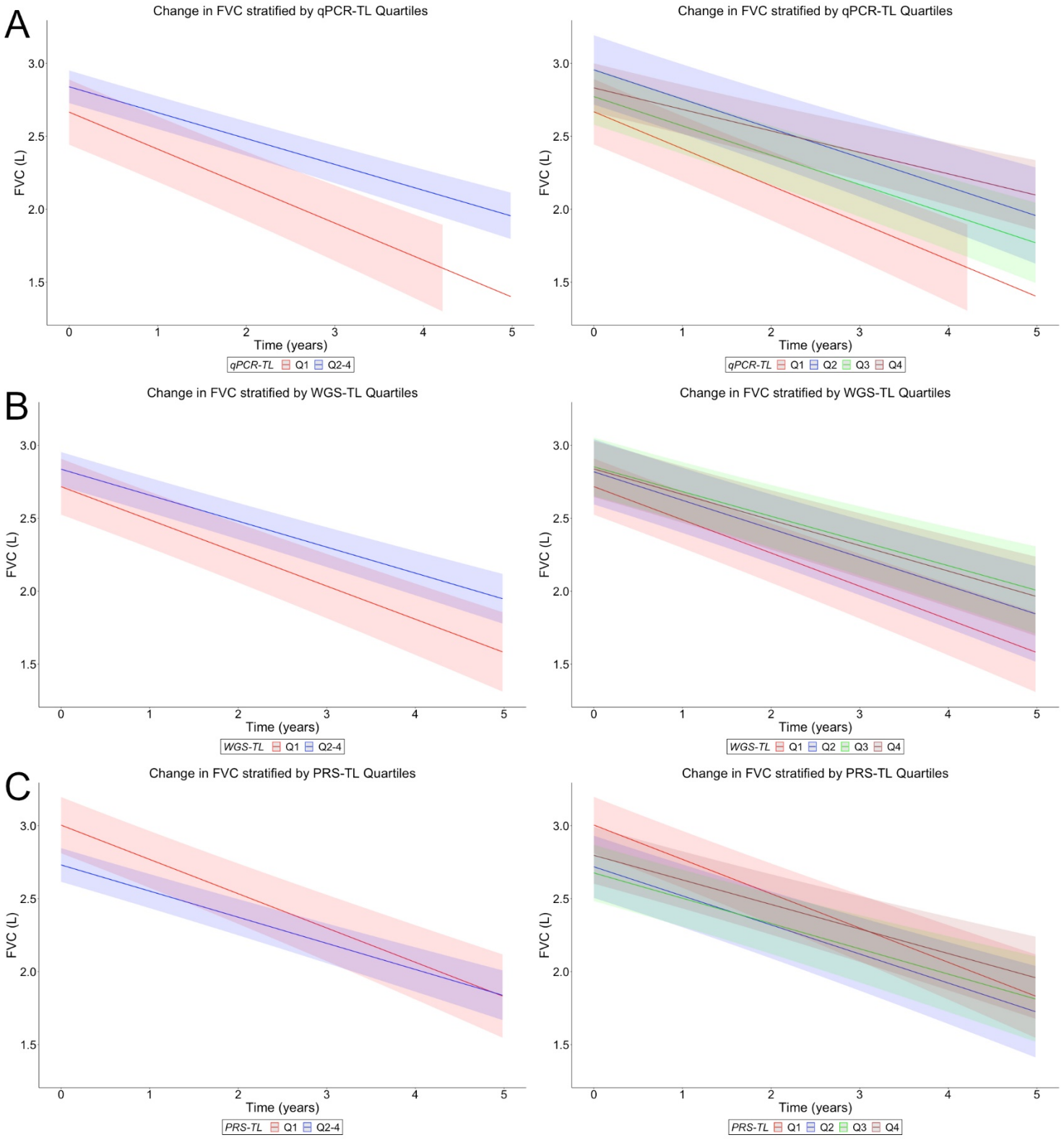


Figure S5. Forced vital capacity decline by genomic factors. The first quartile (Q1) for the WGS-TL and qPCR-TL represents those with the shortest telomere lengths. The first quartile (Q1) for the polygenic risk score represents those genetically predicted to have the shortest telomere lengths.



References

1. Zhang D, Povysil G, Kobeissy PH, Li Q, Wang B, Amelotte MW, et al. Rare and Common Variants in KIF15 Contribute to Genetic Risk of Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2022.
2. Povysil G, Chazara O, Carss KJ, Deevi SVV, Wang Q, Armisen J, et al. Assessing the Role of Rare Genetic Variation in Patients With Heart Failure. JAMA Cardiol. 2021;6(4):379-86.
3. Miller NA, Farrow EG, Gibson M, Willig LK, Twist G, Yoo B, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. Genome Med. 2015;7:100.
4. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80-92.
5. Ren Z, Povysil G, Hostyk JA, Cui H, Bhardwaj N, Goldstein DB. ATAV: a comprehensive platform for population-scale genomic analyses. BMC Bioinformatics. 2021;22(1):149.
6. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet. 2012;91(5):839-48.
7. Cameron-Christie S, Wolock CJ, Groopman E, Petrovski S, Kamalakaran S, Povysil G, et al. Exome-Based Rare-Variant Analyses in CKD. J Am Soc Nephrol. 2019;30(6):1109-22.
8. Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics. 2017;33(17):2776-8.
9. van Moorsel CHM, van der Vis JJ, Grutters JC. Genetic disorders of the surfactant system: focus on adult disease. Eur Respir Rev. 2021;30(159).
10. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.
11. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 2016;99(4):877-85.
12. Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, et al. Predicting the clinical impact of human mutation with deep neural networks. Nat Genet. 2018;50(8):1161-70.
13. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. Nucleic Acids Res. 2020;48(D1):D835-D44.
14. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet. 2017;136(6):665-77.
15. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015;43(Database issue):D789-98.
16. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.
17. Ding Z, Mangino M, Aviv A, Spector T, Durbin R, Consortium UK. Estimating telomere length from whole genome sequence data. Nucleic Acids Res. 2014;42(9):e75.
18. Li C, Stoma S, Lotta LA, Warner S, Albrecht E, Allione A, et al. Genome-wide Association Analysis in Humans Links Nucleotide Metabolism to Leukocyte Telomere Length. Am J Hum Genet. 2020;106(3):389-404.