Variants near TERC are associated with mean telomere length.

Veryan Codd[1,*], **Massimo Mangino**[2,*], **Pim van der Harst**[1,3,*], **Peter S Braund**[1], **Michael Kaiser**[1], **Alan J Beveridge**[1], **Suzanne Rafelt**[1], **Jasbir Moore**[1], **Chris Nelson**[1], **Nicole Soranzo**[2,4], **Guangju Zhai**[2], **Ana M Valdes**[2], **Hannah Blackburn**[2], **Irene Mateo Leach**[3], **Rudolf A de Boer**[3], **Wellcome Trust Case Control Consortium**[5], **Alison H Goodall**[1], **Willem Ouwehand**[6], **Dirk J. van Veldhuisen**[3], **Wiek H. van Gilst**[3], **Gerjan Navis**[7], **Paul R Burton**[8], **Martin D Tobin**[8], **Alistair S Hall**[9], **John R Thompson**[8], **Tim Spector**[2,#], and **Nilesh J Samani**[1,#]

[1]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK [2]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK [3]Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands [4]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK [5]Members of the Wellcome Trust Case Control Consortium are listed in the Supplementary Materials [6]Department of Hematology, University of Cambridge, Cambridge, UK [7]Department of Internal Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands [8]Departments of Health Sciences and Genetics, University of Leicester, Leicester, UK [9]Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, UK

Abstract

We conducted genome-wide association analyses of mean leukocyte telomere length in 2,917 subjects and follow-up replication analyses in 9,492 and identified a locus on 3q26 encompassing the telomerase RNA component *TERC*, with compelling evidence for association (rs12696304, combined P value 3.72×10^{-14}). Each copy of the minor allele of rs12696304 was associated with ≈75 base pairs shorter mean telomere length equivalent to ≈3.6 years of age-related attrition of mean telomere length.

Telomeres are structures at the ends of eukaryotic chromosomes that are made up of a simple repetitive sequence (in humans TTAGGG) and involved in maintaining genomic stability and regulating cellular proliferation.[1] Telomere length (TL) plays an important role in determining telomere function. In somatic cells TL progressively shortens with each mitotic division due to the inability of DNA polymerase to fully replicate the 3 end of the DNA strand. Cellular senescence and subsequent death often occurs when the mean TL

reaches a critical value.[1] Shorter mean leukocyte TL has been shown to be associated with risk of several age-related diseases and proposed as a marker of biological ageing. Maintenance of TL is required in certain cell types (e.g. germ cells) and this is achieved by telomerase, a ribonucleoprotein consisting of a reverse transcriptase (TERT) and a RNA template (TERC) that facilitates addition of the telomere repeat sequence. Telomerase reactivation and TL maintenance also contribute to the pathogenesis of several cancers.[1]

TL has a strong genetic determination with heritability estimates ranging from 44-80%.[2,3] Quantitative trait linkage (QTL) studies have mapped putative loci for TL to human chromosomes 3p26.1, 10q26.13, 12q12.22 and 14q23.2.[3-5] A recent genome-wide association (GWA) study identified two SNPs on chromosome 18q12.2 associated with TL, although not at a genome-wide significant level.[6] To identify additional variants that affect TL we undertook GWA analyses in two large cohorts of European descent followed by replication of promising signals in three further European sample sets.

The discovery cohorts comprised 1,487 individuals with coronary artery disease (CAD) from the British Heart Foundation Family Heart Study (BHF-FHS),[7] and 1,430 United Kingdom Blood Service donors (UKBS)[8] that had genome-wide SNP genotype data generated using the Affymetrix 500K array as part of the WTCCC Study.[8] Further details of the cohorts and genotyping are given in Supplementary Methods and Supplementary Table 1. Mean leukocyte TL was measured using a quantitative PCR-based technique,[9] which expresses mean TL as a ratio (T/S) of telomere repeat length (T) to copy number of a single copy gene, 36B4(S), within each sample (see Supplementary Methods). The T/S ratios were distributed normally (Supplementary Figures 1A) and showed the expected age-related attrition in TL (Supplementary Figure 1B) in both cohorts.

We analysed the association of T/S ratio adjusted for age and gender with genotype individually in the BHF-FHS (Supplementary Table 2A) and UKBS (Supplementary Tables 2B) cohorts and also in a combined analysis of the two cohorts. The quantile-quantile plots for each cohort are shown in Supplementary Figure 1C and the power to detect associations in Supplementary Figure 1D. The genomic inflation control factors for the BHF-FHS and UKBS analyses were 1.02 and 0.99, respectively. We screened the results from the combined analysis for SNPs that showed concordance in their results in the individual analyses (i.e. at least nominally significant (P<0.05) in both cohorts with beta coefficients in the same direction). These criteria identified 180 SNPs at $p<1\times10^{-3}$ and 24 SNPs at $p<1\times10^{-4}$. SNPs achieving a combined $P<5\times10^{-5}$ are shown in Supplementary Table 2C. Notably, neither the previously reported SNPs on chromosome 18q12.2[6] nor SNPs at any of the loci identified by QTL analysis[3-5] showed a consistent association with TL in our cohorts (Supplementary Table 3A). We set a pragmatic threshold of $P<1\times10^{-5}$ in the combined analysis to take SNPs forward for replication. This identified two SNPs: rs610160 (Chr 11q22), and rs10511887 (Chr 9p21.1) (Table 1). However one SNP, rs16847897 (Chr 3q26), which fell marginally outside this threshold (combined $P=1.03\times10^{-5}$) was the lead SNP from a locus that includes the *TERC* sequence. A further SNP in the locus, rs12696304, with a relatively weak linkage disequilibrium (LD) ($r^2=0.49$) to rs16847897 also showed moderate association (combined $P=9.33\times10^{-5}$). Due to the presence of a candidate gene at this locus these SNPs were also taken to replication (Table 1). None of the SNPs taken forward showed any association with CAD.

Details of the replication cohorts are given in Supplementary Methods and Supplementary Table 1. Initial replication was performed in 2020 subjects of the GRAPHIC study.[10] Leukocyte mean TL was determined in these subjects using the same PCR technique used in the discovery cohorts. Neither rs610160 nor rs10511887 showed any evidence of replication. However, both SNPs on chromosome 3q26 showed significant associations in

the same direction as the discovery cohorts (Table 1). Next, we analysed data on leukocyte mean TL and genotypes for rs16847897 and rs12696304 obtained in 3256 subjects from the TwinsUK study.[11]Mean TL in this cohort was measured by Southern blotting (see Supplementary Methods). Despite the different method of measuring TL, there was again a significant and similar association of both SNPs with TL (Table 1, and Supplementary Table 3B). Finally, we examined the association of the 3q26 SNPs with TL in 4216 individuals from the PREVEND study.[12] TL was measured in these subjects using a variation of the PCR-based method used in the discovery and GRAPHIC cohorts (see Supplementary Methods). Both rs16847897 and rs12696304 again showed significant associations with TL (Table 1). To assess the totality of the evidence for the association of the chromosome 3q26 locus with TL, we used Fisher's method (see Supplementary Methods). We also Z-transformed the individual TL measurements in each study to obtain comparable results (Supplementary Table 4A) and meta-analysed the results using METAL (see Supplementary Methods). This gave P values of 2.79×10^{-12} and 3.72×10^{-14} for rs16847897 and rs12696304, respectively.

Figure 1 shows the association signal across the 3q26 locus from the combined BHF-FHS and UKBS analysis. rs12696304 and rs16847897 are actually located towards opposite ends of an ≈ 87kb region showing significant association. Interestingly, despite the relatively weak LD between rs12696304 and rs16847897, conditional analysis showed that their associations are not independent. Inclusion of rs16847897 as a covariate in the association analysis of rs12696304 abolished the latter's association with TL (P=0.40) and vice versa (P=0.10). Haplotype analysis of SNPs at the locus also suggested a single signal (see Supplementary Methods and Supplementary Table 4B).

The variance in TL explained by the 3q26 locus ranged from 0.32% in PREVEND to 1.0% in BHF-FHS for rs12696304. The TwinsUK study provides a direct estimate in base pairs (bp) of the effect of the locus on TL. Each minor allele of rs12696304 was associated with a mean leukocyte TL that was ≈75 bp shorter in this cohort (Table 1). To put this in context, in cross-sectional analysis, TL decreased by ≈21bp per year greater age in the TwinsUK cohort; therefore each minor allele of rs12696304 was associated with shorter mean TL equivalent to ≈3.6 years of average age-related telomere attrition. Similar analyses of the combined BHF-FHS and UKBS data and the PREVEND study gave effect sizes per minor allele of rs12696304 equivalent to 4.0 years and 6.6 years of age-related attrition, respectively.

rs12696304 lies 1.5kb downstream of *TERC* (Figure 1). To investigate the possibility that the genotyped variants are markers for causal variants within *TERC* that affect its function, we sequenced the coding region of *TERC* (451bp) and approximately 1kb upstream and 1.5kb downstream in individuals from the GRAPHIC study that were homozygous for either both minor alleles (n=16) or both major alleles (n=16) of rs12696304 and rs16847897 (see Supplementary Methods). No variants were identified within the *TERC* coding sequence. However, this does not exclude the possibility that the association is mediated by an effect on *TERC* expression. Alternately, the effect on TL could be through one of the other genes in the 3q26 locus. These include *ARPM1* (Actin-related protein M1), *MYNN* (myoneurin) and three members of the leucine-rich repeat containing (*LRRC*) superfamily (*LRRC34, LRRC31 and LRRIQ4*). *ARPM1* is a nuclear protein thought to have a role in organisation of the sperm-specific nucleus in the mouse.[13] Although very little is directly known about *LRRC34, LRRC31 and LRRIQ4,* the LRRC superfamily consists of members with diverse roles, including DNA repair, cell-cycle regulation, apoptosis and chromosomal stability.[14,15]

In summary, we report a locus on 3q26 that affects telomere length in humans. Given the importance of telomeres in nuclear and cellular function and the central role of telomere

length in determining telomere function, our finding could have broad relevance for both normal and pathological age-associated processes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

1. Blackburn EH, Greider CW, Szostak JW. Nat. Med. 2006; 12:1133–1138. [PubMed: 17024208]

2. Njajou OT. PNAS. 2007; 104(29):12135–9. [PubMed: 17623782]

3. Vasa Nicotera M, et al. Am. J. Hum. Genet. 2005; 76:147–51. [PubMed: 15520935]

4. Andrew T, et al. Am. J. Hum. Genet. 2006; 78:480–6. [PubMed: 16400618]

5. Mangino M, et al. Hum. Mol. Genet. 2008; 17:2518–23. [PubMed: 18487243]

6. Mangino M, et al. J. Med. Genet. 2009; 46(7):451–4. [PubMed: 19359265]

7. BHF Family Heart Study Research Group. Am. J. Hum. Genet. 2005; 77:1011–1020. [PubMed: 16380912]

8. Wellcome Trust Case Control Consortium. Nature. 2007; 447:661–78. [PubMed: 17554300]

9. Cawthon RM. Nucleic Acids Res. 2002; 30:e47. [PubMed: 12000852]

10. Tobin MD, et al. Hypertension. 2008; 51:1658–1664. [PubMed: 18443236]

11. Spector TD, Williams FM. Twin. Res. Hum. Genet. 2006; 9:899–906. [PubMed: 17254428]

12. Hillege HL, et al. Circulation. 2002; 106:1777–82. [PubMed: 12356629]

13. Hara Y, et al. FEBS Lett. 2008; 582:2998–3004. [PubMed: 18692047]

14. Kobe B, Kajava AV. Curr. Opin. Struct. Biol. 2001; 11:725–732. [PubMed: 11751054]

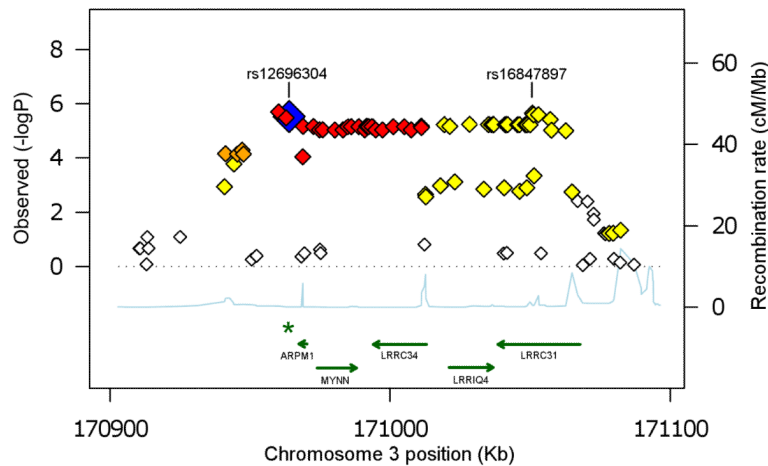15. Rabenau, et al. Oncogene. 2004; 23:5056–5067. [PubMed: 15107827]

Figure 1.

A map of the chromosome 3q26 region showing genetic association with mean leukocyte telomere length.

Association of individual SNPs from the combined BHF-FHS and UKBS analysis is plotted as $-\log_{10}P$ against chromosomal base pair position. Results of both genotyped and imputed SNPs are shown. rs12696304 is shown in blue and the LD relationship of the other markers (including rs16847897) to rs12696304 SNP is indicated by colour: red $R^2>0.8$, orange $R^2>0.5$, yellow $R^2>0.2$, white $R^2<0.2$. The location of known genes in the region *ARPM1* (Actin-related protein M1), *MYNN* (myoneurin) and three members of the leucine-rich repeat containing (*LRRC*) superfamily (*LRRC34, LRRC31 and LRRIQ4*) are shown as well as the location of recombination hot spots (blue peaks). The position of the *TERC* coding sequence (451bp) is represented by *

Table 1

Telomere length association results from the GWAS studies and in replication studies for the 3q26.2, 9p21.1 and 11q22.3 loci

Chr	SNP	Minor Allele	BHF-FHS GWAS			UKBS GWAS			BHF-FHS+UKBS			GRAPHIC			TwinsUK			PREVEND		
			MAF	Beta (SE)	Pvalue	MAF	Beta (SE)	Pvalue	MAF	Beta (SE)	Pvalue	MAF	Beta (SE)	Pvalue	MAF	Beta (SE)	Pvalue	MAF	Beta (SE)	Pvalue
3	rs12696304	G	0.26	-0.039 (0.009)	9.73×10^{-5}	0.26	-0.022 (0.013)	8.34×10^{-2}	0.26	-0.030 (0.008)	9.33×10^{-5}	0.26	-0.028 (0.009)	1.39×10^{-3}	0.25	-0.075 (0.019)	8.13×10^{-5}	0.30	-0.030 (0.008)	3.21×10^{-4}
3	rs16847897	C	0.26	-0.033 (0.009)	1.25×10^{-4}	0.27	-0.035 (0.013)	6.64×10^{-3}	0.26	-0.034 (0.008)	1.03×10^{-5}	0.26	-0.029 (0.009)	9.34×10^{-4}	0.26	-0.064 (0.019)	6.44×10^{-4}	0.30	-0.024 (0.009)	6.40×10^{-3}
9	rs10511887	G	0.24	-0.037 (0.011)	7.41×10^{-4}	0.24	-0.052 (0.017)	2.10×10^{-3}	0.24	-0.044 (0.010)	8.86×10^{-6}	0.25	-0.001 (0.009)	0.91		–			–	
11	rs610160	C	0.13	0.030 (0.011)	8.89×10^{-3}	0.12	0.064 (0.017)	2.03×10^{-4}	0.13	0.046 (0.010)	7.05×10^{-6}	0.11	-0.007 (0.013)	0.57		–			–	

The Table shows the association findings with mean leukocyte telomere length of the 4 SNPs identified in the combined analyses of the British Heart Foundation Family Heart study (BHF-FHS) and the United Kingdom Blood Service (UKBS) donors genome wide association studies (GWAS) and taken through to replication in the GRAPHIC study and subsequently for rs12696304 and rs16847897 into the TwinsUK and PREVEND studies. The minor allele frequency (MAF), beta coefficient (Beta), standard error (SE) and the P value for association for each cohort individually and for the combined analysis of the BHF-FHS and UKBS subjects are shown. SNP associations were analysed using an additive model. The beta coefficients equate to the mean change in telomere length for the particular measure of telomere length (T/S ratio in BHF-FHS, UKBS, GRAPHIC and PREVEND and Southern blot determined telomere length in TwinsUK) per copy of the minor allele for the specific SNP. A negative beta indicates a shorter mean telomere in those carrying the minor allele.