



OPEN ACCESS

ORIGINAL ARTICLE

# Genome-wide association study of telomere length among South Asians identifies a second RTEL1 association signal

Dayana A Delgado,<sup>1</sup> Chenan Zhang,<sup>1,2</sup> Lin S Chen,<sup>1</sup> Jianjun Gao,<sup>3</sup> Shantanu Roy,<sup>1,4</sup> Justin Shinkle,<sup>1</sup> Mekala Sabarinathan,<sup>1</sup> Maria Argos,<sup>5</sup> Lin Tong,<sup>1</sup> Alauddin Ahmed,<sup>6</sup> Tariqul Islam,<sup>6</sup> Muhammad Rakibuz-Zaman,<sup>6</sup> Golam Sarwar,<sup>6</sup> Hasan Shahriar,<sup>6</sup> Mahfuzar Rahman,<sup>7</sup> Mohammad Yunus,<sup>8</sup> Farzana Jasmine,<sup>1</sup> Muhammad G Kibriya,<sup>1</sup> Habibul Ahsan,<sup>1,9,10,11</sup> Brandon L Pierce<sup>1,9,10</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2017-104922>).

For numbered affiliations see end of article.

## Correspondence to

Dr Brandon L Pierce, Department of Public Health Sciences, The University of Chicago, 5841 South Maryland Avenue, MC2000 Chicago, IL 60637, USA; [brandonpierce@uchicago.edu](mailto:brandonpierce@uchicago.edu)

Received 13 July 2017  
Revised 13 September 2017  
Accepted 26 September 2017  
Published Online First  
18 November 2017

## ABSTRACT

**Background** Leucocyte telomere length (TL) is a potential biomarker of ageing and risk for age-related disease. Leucocyte TL is heritable and shows substantial differences by race/ethnicity. Recent genome-wide association studies (GWAS) report ~10 loci harbouring SNPs associated with leucocyte TL, but these studies focus primarily on populations of European ancestry.

**Objective** This study aims to enhance our understanding of genetic determinants of TL across populations.

**Methods** We performed a GWAS of TL using data on 5075 Bangladeshi adults. We measured TL using one of two technologies (qPCR or a Luminex-based method) and used standardised variables as TL phenotypes.

**Results** Our results replicate previously reported associations in the *TERC* and *TERT* regions ( $P=2.2\times 10^{-8}$  and  $P=6.4\times 10^{-6}$ , respectively). We observed a novel association signal in the *RTEL1* gene (intronic SNP rs2297439;  $P=2.82\times 10^{-7}$ ) that is independent of previously reported TL-associated SNPs in this region. The minor allele for rs2297439 is common in South Asian populations ( $\geq 0.25$ ) but at lower frequencies in other populations (eg, 0.07 in Northern Europeans). Among the eight other previously reported association signals, all were directionally consistent with our study, but only rs8105767 (*ZNF208*) was nominally significant ( $P=0.003$ ). SNP-based heritability estimates were as high as 44% when analysing close relatives but much lower when analysing distant relatives only.

**Conclusions** In this first GWAS of TL in a South Asian population, we replicate some, but not all, of the loci reported in prior GWAS of individuals of European ancestry, and we identify a novel second association signal at the *RTEL1* locus.

## INTRODUCTION

Telomeres are DNA–protein complexes that protect chromosome ends.<sup>1</sup> In differentiated cells, the progressive attrition of the DNA component of telomeres, a tandem repeat sequence (TTAGGG)<sub>n</sub>, with each round of mitotic division ultimately leads to loss of telomere protection, triggering cell senescence or apoptosis.<sup>1,2</sup> In germ cells, stem cells and most cancer cells, synthesis and maintenance of

telomeres is carried out by the holoenzyme, telomerase, a ribonucleic protein consisting of a reverse transcriptase (*TERT*) and RNA template (*TERC*), which enable the elongation of the telomere repeat sequence, conferring replicative immortality.<sup>3–5</sup> Leucocyte telomere length (TL) shortens with age, and associations between short TL and increased risk for numerous age-related diseases have been reported (ie, cardiovascular disease, hypertension, liver disorders, diabetes, atherosclerosis and mortality); consequently, TL is regarded as a surrogate marker for biological ageing.<sup>6–10</sup> Additionally, there is increasing evidence suggesting more complex relationships among TL, cellular senescence and human health, with longer TL conferring increased risk for several cancer types, including melanoma,<sup>11</sup> glioma,<sup>12</sup> neuroblastoma,<sup>13</sup> lung adenocarcinoma<sup>14</sup> and chronic lymphocytic leukaemia.<sup>15</sup>

Leucocyte TL is influenced by inherited genetic factors.<sup>16–19</sup> Twin studies report TL heritability estimates as high as 80%,<sup>19</sup> while other family studies estimate a heritability ranging between 34% and 50%.<sup>18</sup> Genome-wide association studies (GWAS) have shown that leucocyte TL is influenced by inherited genetic polymorphisms. These studies have identified loci harbouring common genetic variants associated with leucocyte TL, including: *ACYP2* (2p16.2), *TERC* (3p26), *TERT* (5p15.33), *NAF1* (10q24.33), *OBFC1* (10q24.33), *DCAF4* (14q24.2), *CTC1* (17p13.1), *ZNF208* (19p12), *ZNF676* (19p12) and *RTEL1* (20q13.3).<sup>16–29</sup> These variants have been discovered and replicated in studies of individuals of primarily European ancestry. Considering the strong evidence of population differences in TL, which may be due in part to polygenic adaptation,<sup>30</sup> genetic studies of TL in non-European populations are warranted.<sup>31</sup> In this study, we undertook S of TL among Bangladeshi individuals to identify new TL-associated variants and confirm the effects of loci reported in prior studies.

## EXPERIMENTAL PROCEDURES

### Study participants

The Health Effects of Arsenic Longitudinal Study (HEALS) is a prospective cohort study in Araihaaz, Bangladesh, designed to assess the health effects of



CrossMark

**To cite:** Delgado DA, Zhang C, Chen LS, et al. *J Med Genet* 2018;**55**:64–71.

arsenic exposure through drinking water. Details of the HEALS have been described previously.<sup>32</sup> In brief, 11 746 participants of ages 18–75 years old were recruited between October 2000 and May 2002. An expansion of the HEALS cohort (ACE) by 8287 participants occurred between 2006 and 2008. The Bangladesh Vitamin E and Selenium Trial (BEST) is a 2×2 factorial, double-blind, randomised chemoprevention trial designed to assess the effect of vitamin E and selenium supplementation on non-melanoma skin cancer risk among 7000 adults. BEST participants were recruited from Araihasar, Matlab and other surrounding districts of Bangladesh. Details of the BEST study methods have been described previously.<sup>33</sup>

### DNA extraction and genotyping

Blood collection, DNA extraction, genotyping and quality control have been described previously.<sup>34,35</sup> To summarise, in each study, study physicians collected blood from all participants at baseline. Genomic DNA for HEALS samples was extracted from clot blood using the Flexigene DNA kit (Cat #51204) from Qiagen (Valencia, California, USA). For BEST samples, DNA was extracted from whole blood using the QIAamp 96 DNA Blood Kit (Cat #51161) from Qiagen. A total of 5499 DNA samples were genotyped on HumanCytoSNP-12 v2.1 chips with 299 140 SNPs measured. Genotyping quality control was performed using PLINK. Individuals with call rates of <97% (n=13), gender mismatches (n=79) and technical replicate samples or duplicates (n=53) were removed. SNPs that had low call rates, that is (<95%) (n=20), were monomorphic (n=39 798), did not have rs identifiers (n=941) and had hardy-weinberg equilibrium (HWE) P values of <10<sup>-10</sup> (n=634), were excluded. This quality control resulted in 5354 individuals and 257 747 genotyped SNPs. We generated imputed genotypes using the 1000 Genomes Project reference panel. We excluded SNPs that had an imputation r<sup>2</sup><0.3, were triallelic, had an HWE of P<10<sup>-10</sup> and had minor allele frequency (MAF) ≤0.01. HWE testing in an unrelated set of participants confirms that no SNPs used in the analysis have an HWE P value of <10<sup>-7</sup>.

### Measurement of TL using qPCR

We attempted TL measurement for all participants having existing genome-wide SNP data. qPCR was performed on 2203 HEALS samples using two different plate designs (plate designs 1 and 2) on 96-well plates in order to measure each subject's telomere repeat abundance as well as the abundance of a two-copy reference gene (*RPLP0*). Reference samples were included on each plate to allow standardisation of each reaction run with respect to interplate variations. For plate design 1, each plate consisted of triplicates of 14 participant DNA samples and six replicates of a reference DNA sample, each measured for both telomere and single copy gene for a total of 96 samples per plate. The 14 possible positions and corresponding primers for each plate using this design are described in online supplementary figure S1. For plate design 2, telomere and single copy gene amplification of each DNA sample were performed on paired plates, with the sample placed at the same position on each of the two plates. Each plate consisted of triplicates of 31 participant DNA samples and triplicates of a reference DNA sample, each measured for either telomere or single copy gene at the same position on paired plates. The 31 possible positions for each plate using this design are described in online supplementary figure S2.

In brief, both plate designs used 30 μL amplification reaction volume of 1× Qiagen Quantitect Sybr Green Master Mix

(Qiagen) and 25 ng of template DNA. Both plate designs used the following primer sequences: 5'GGTTTTTGGGGTGAGGGT GAGGGTGAGGGTGAGGGT3' (forward), 5'TCCCCACTATC CCTATCCCTATCCCTATCCCTATCCCTA3' (reverse) for telomere sequence amplification and 5'CAGCAAGTGGGAAGG TGTAATCC3' (forward), 5'CCCATTCTATCATCAACGG GTACAA3' (reverse) for 36B4 sequence amplification.

For plate design 1, we used a protocol previously described by Ehrlenbach *et al.*<sup>36</sup> The qPCR reaction was performed under the following conditions: 95°C for 15 min to activate Taq polymerase, 30 cycles at 95°C for 15 s, 54°C for 2 min and 72°C for 1 min, followed by 72°C for 4 min. For plate design 2, we used a protocol previously described by Cawthon.<sup>37</sup> The telomere reaction used the following protocol: 95°C for 15 min to activate Taq polymerase, 25 cycles at 95°C for 15 s and 54°C for 2 min, while the *RPLP0* (reference) reaction ran at 95°C for 15 min to activate Taq polymerase, 30 cycles at 95°C for 15 mins and 58°C for 1 min.

Threshold cycle (C<sub>t</sub>), is the threshold that represents a level above background (noise) fluorescence and is the number of cycles necessary to detect a real signal from the sample. C<sub>t</sub> values were averaged across replicates for each sample. The relative TL was calculated as a ratio of telomere repeat copy number to single-gene (*RPLP0*) copy number, that is, T/S ratio (telomere to single gene reference), according to the following equation:

$$T/S = \frac{2^{C_{t\text{telomere}}(\text{Reference})/2^{C_{t\text{telomere}}(\text{Sample})}}{2^{C_{t\text{RPLP0}}(\text{Reference})/2^{C_{t\text{RPLP0}}(\text{Sample})}} \quad (1)$$

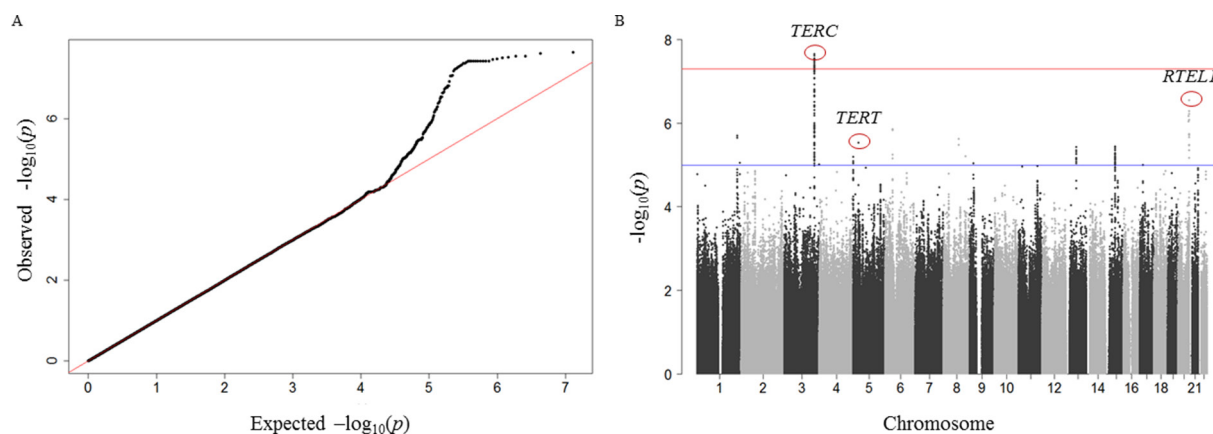
The coefficients of variation (CV) were calculated for each set of replicates (one for each study sample triplicate and one for each reference sample sextuplicate). The CV was calculated as the SD of the sample replicates divided by the mean of the replicates:

$$CV = 100 \times \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \div \bar{x} \quad (2)$$

To confirm reproducibility of the measured T/S ratio for plate design 1, we reran 37 samples on separate days, from which we calculated the interplate CV to be 11.7%. An additional sample across 31 plates assessed for reproducibility of plate design 2 showed an interplate CV of 9.8% in the assessment of reproducibility of plate design 2.

### Measurement of TL using QuantiGene chemistry on a Luminex platform

The Affymetric-Panomics QuantiGene Plex, a probe-based non-PCR assay, was developed in our lab to measure TL in BEST (n=1825) and ACE (n=1047). This assay uses approximately 50 nanograms of DNA which is hybridised to sequence-specific probes for the telomere repeat sequence specific and the reference gene sequence (*ALK*). The abundance of the telomere repeat sequence, relative to the reference gene, is detected using a Luminex instrument. The details of this method have been previously described.<sup>38</sup> A major advantage of the Luminex-based method over qPCR is the lack of position effects.<sup>39</sup> Additionally, the Luminex method has been validated by Pierce *et al.*<sup>40</sup> in a blinded comparison with Southern blot analysis. Briefly, fluorescent Luminex microbeads designed with capture probes are exploited to capture DNA molecules. Each DNA target sequence has two sets of specific probes: (1) capture extenders and (2) label extenders and blocker probes. The intensities of the fluorescent signals produced from the binding of preamplifiers to amplifiers are read on a Luminex 200 instrument and are proportional to the relative TL expressed as Telomere Quantity



**Figure 1** Summary plots of GWAS results. (A) Quantile–quantile (Q–Q) plot of the negative logarithm of the observed (y-axis) and expected (x-axis) P value for each SNP, where the red line represents the null hypothesis of no associations. (B) Manhattan plot showing association signals in *TERC*, *TERT* and *RTEL1* regions. The  $-\log_{10}(P)$  values are plotted against physical position for 6.3 million SNPs. The red line indicates the genome-wide significance threshold at  $P=5\times 10^{-8}$ . We used a threshold of  $P=1\times 10^{-5}$  to define suggestive signals (blue line). GWAS, genome-wide association study.

Index (TQI). Similar to T/S ratio obtained through qPCR, TQI is a ratio of telomere probe to reference gene probe that directly accounts for plate-to-plate variation in intensity measurement.

### Statistical analyses

We used linear mixed-effects models to remove experimental variation in our TL measures; categorical indicator variables representing position and plate were treated as fixed and random effects, respectively, using the lme package in R (see online supplementary figure S3). We ran three mixed-effects models; one for each different qPCR plate design used and one for the Luminex-based method. The residuals from these mixed-effects models were standardised and used as our TL phenotype in the GWAS. To ensure adequate control for population structure (ie, cryptic relatedness), we ran principal component analysis using the 1000 Genomes Project reference populations ( $n=2504$ ) and unrelated ( $r<0.05$ ) participants in our Bangladeshi cohort ( $n=1524$ ) (see online supplementary figure S4). Because of the presence of close relatives in these cohorts (see online supplementary table S1), we use the Genome-wide Complex Trait Analysis (GCTA) software<sup>41</sup> for the GWAS using a mixed linear model that adjusts for relatedness and the following fixed covariates: age, sex, batch and cohort.<sup>42</sup> For each region identified in the GWAS, we conducted a conditional association analysis, adjusting for the lead SNP, to identify secondary association signals. Observed association signals were compared with summary results from the European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium GWAS of individuals of European ancestry provided by Codd *et al*<sup>21</sup> to test for replication (<https://downloads.lcbru.le.ac.uk/engage>).

In order to obtain a SNP-based heritability estimate for TL, we used a restricted maximum likelihood analysis as implemented in

GCTA, adjusting for cohort, sex, batch and age. We conducted these analyses first restricting to ‘unrelated’ individuals with coefficients of relationship ( $r$ ) of 0.05 and below. We then increased this threshold to include more closely related pairs, which increases power but is prone to bias due to shared environment. In addition, we used the method described by Zaitlen *et al*<sup>43</sup> to estimate the full narrow-sense  $h^2$ , a method that leverages relatedness for close relatives only by setting  $r$  values less than 0.05 to zero. This method captures the effects of rare variants but is also prone to bias due to shared environment.

### RESULTS

Characteristics of the 5075 participants included in our analyses are described in online supplementary table S2. Similar to other studies of TL, we observed a significant inverse association between TL and age that is consistent by cohort (HEALS:  $r=-0.20$ ,  $P<2\times 10^{-16}$ , BEST:  $r=-0.23$ ,  $P<2\times 10^{-16}$  and ACE:  $r=-0.28$ ,  $P<2\times 10^{-16}$ ). The results of the GWAS are summarised in figure 1. The quantile–quantile plot shows no evidence of early deviation from the P value distribution under the null hypothesis, suggesting adequate control for population structure (ie, cryptic relatedness) and potential batch effects (genomic inflation factor=0.999) (figure 1A and online supplementary figure S4). Adjustment for body mass index and smoking status did not impact the findings of this GWAS. Among our eight most significant association signals, three reside in regions previously reported to affect leucocyte TL, and each of these three regions contains a known telomere maintenance gene: 3q26 (*TERC*), 5p15.33 (*TERT*) and 20p13.3 (*RTEL1*) (table 1 and figure 1A).

The *TERC* region (3q26) was the only region containing SNPs passing a genome-wide significance threshold of

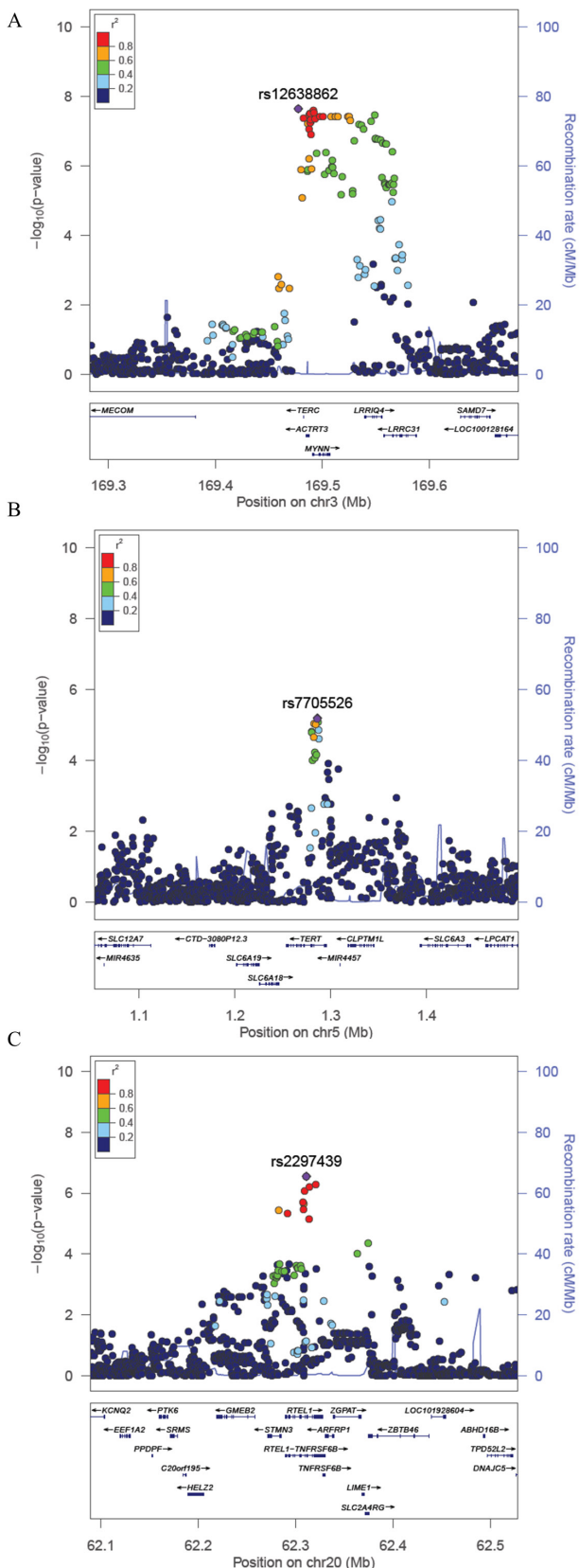
**Table 1** Associations\* observed in regions previously reported to affect telomere length (TL) based on a genome-wide association analysis of 6.3 million SNPs ( $n=5075$ )

SNP identifier	Chr	Gene	Major allele	Minor allele	MAF	$\beta$ †	SE	P
rs12638862	3	<i>TERC</i>	A	G	0.24	−0.13	0.02	$2.2\times 10^{-8}$
rs7705526	5	<i>TERT</i>	C	A	0.43	0.09	0.02	$6.4\times 10^{-6}$
rs2297439	20	<i>RTEL1</i>	T	G	0.25	−0.12	0.02	$2.8\times 10^{-7}$

\*Association estimates are adjusted for sex, age, TL measurement type and measurement batch.

†Association estimates are for the minor allele. The major allele is the reference allele.

MAF, minor allele frequency.



**Figure 2** Regional association plots for loci associated with TL. (A–C) The  $-\log_{10}(P \text{ Value})$  for each SNP is plotted against the base-pair position along each chromosome (Mb). For each locus, the lead SNP is represented in purple; SNPs are colour coded by level of linkage disequilibrium ( $r^2$ ) to the lead SNP, and the blue lines represent recombination rates (cM/Mb). TL, telomere length.

$P < 5 \times 10^{-8}$  (table 1). The SNP showing the strongest association was rs12638862 ( $P = 2.2 \times 10^{-8}$ ), with the minor allele G (MAF=0.24) associated with decreased TL ( $\beta = -0.13$ ) (table 1, figure 2A). While this SNP (rs12638862) has not been reported as a lead SNP in prior GWAS of TL, it is in linkage disequilibrium (LD) with previously reported SNPs ( $r^2 = 0.87$ , 0.84 and 0.54 for rs10936599, rs3772190, and rs12696304, respectively, based on 1000 Genomes Project Bangladesh (BEB) data). Conditional analysis adjusting for rs12638862 showed no evidence of a secondary association signal in this region (see online supplementary figure S5a), which is consistent with prior findings.<sup>21</sup>

The SNP in the *TERT* region showing the strongest association was rs7705526 ( $P = 6.4 \times 10^{-6}$ ), whose minor allele A (MAF=0.43) was associated with decreased TL ( $\beta = -0.09$ ) (table 1, figure 2B). Our lead *TERT* SNP was the same lead SNP reported by Do *et al*<sup>22</sup> and was in LD ( $r^2 = 0.57$ ) with the lead SNP reported in Codd *et al*<sup>21</sup> (rs2736100). The direction of association was also consistent with prior studies, and conditional analyses adjusted for the top SNP showed no evidence of a secondary association signal (see online supplementary figure S5b).

In the *RTEL1* region, our lead SNP was rs2297439 ( $P = 2.8 \times 10^{-7}$ ); the minor allele G (MAF=0.25) was associated with decreased TL ( $\beta = -0.12$ ) (table 1, figure 2C). In contrast to the association in the *RTEL1* region previously reported by Codd *et al*<sup>21</sup> (lead SNP rs755017 in exon 2 of *ZBTB46*), our lead SNP is located in intron 12 of the *RTEL1* gene (>100 kb from rs755017) and is not in LD with rs755017 ( $r^2 = 0.06$  and  $r^2 = 0.01$  based on 1000 Genomes Populations BEB and European (EUR), respectively), suggesting that our lead SNP is tagging a different causal variant in the *RTEL1* region (see Discussion and online supplementary figure S6). Furthermore, this SNP was not reported in the summary results provided by Codd *et al*<sup>21</sup> potentially because it has a low MAF in populations of North and Western European ancestry (MAF=0.07 in Utah Residents with Northern and Western European ancestry (CEU), Finnish in Finland (FIN), and British in England and Scotland (GBR) 1000 Genomes groups) and may be difficult to accurately impute. In fact, this variant has a low MAF in all 1000 Genomes populations except for those of South Asian ancestry (see online supplementary table S3 and figure S7). Based on the 1000 Genomes BEB population, there are three variants in perfect LD with rs2297439 ( $r^2 = 1$ ) (rs77552606, rs114703330 and rs75031349) and four additional variants in high LD ( $r^2 > 0.9$ ) (rs78430218, rs79981841, rs16983884 and rs79981941), all of which also show a low MAF in populations of North and Western European ancestry (see online supplementary table S4). Conditional analyses of this region adjusted for the top SNP showed no evidence of a secondary association signal (see online supplementary figure S5c). The previously reported *RTEL1* SNP rs755071 showed little evidence of association with TL in our Bangladeshi population ( $P = 0.92$ ), and nor did SNPs in LD with rs755071 (see online supplementary figure S8). The association we observed between TL and our lead SNP in the *RTEL1* locus is consistent by age, sex and cohort.

In addition to *TERC*, *TERT* and *RTEL1*, we observed five additional loci with SNPs passing a P value threshold of  $1 \times 10^{-5}$ . These include 1q42.1 (rs1151814,  $P = 2.0 \times 10^{-6}$ ), 6p21.1 (rs9357354,  $P = 1.4 \times 10^{-6}$ ), 8p12 (rs75283006,  $P = 1.7 \times 10^{-5}$ ), 13q21.1 (rs9537514,  $P = 4.4 \times 10^{-6}$ ) and 15p11.15 (rs28790308,  $P = 3.6 \times 10^{-6}$ ) (table 2 and online supplementary figure S9). Only three of these five SNPs were tested in Codd *et al*<sup>21</sup> (1q42.1 (rs1151814), 6p21.1 (rs9357354) and 13q21.1 (rs9537514)), but none of these three were replicated in individuals of European ancestry based on the results from

**Table 2** Suggestive association signals ( $P < 1 \times 10^{-5}$ ) observed in a genome-wide association study (GWAS) of 6.3 million SNPs, 2000–2009 ( $n=5075$ )\*

SNP identifier	Chr	Closest gene	Major allele	Minor allele	MAF	$\beta$ †	SE	P
rs1151814	1	<i>PARP1</i>	T	C	0.10	-0.16	0.03	$2.0 \times 10^{-6}$
rs9357354	6	<i>NCR2</i>	G	C	0.17	0.13	0.03	$1.4 \times 10^{-6}$
rs75283006	8	<i>RBPMS-AS1</i>	A	G	0.09	0.15	0.03	$1.7 \times 10^{-5}$
rs9537514	13	<i>PRR20</i>	G	T	0.21	0.11	0.02	$4.4 \times 10^{-6}$
rs28790308	15	<i>ONECUT1</i>	C	G	0.25	-0.10	0.02	$3.6 \times 10^{-6}$

\*GWAS was adjusted for sex, age, TL measurement type and batch effects.

†Association estimates are for the minor allele. The major allele is the reference allele.

GWAS, genome-wide association studies; MAF, minor allele frequency; TL, telomere length.

the ENGAGE consortium (see online supplementary table S5), although the MAFs for these SNPs are similar in European and South Asian populations.

We tested all nine SNPs previously reported to be associated with TL in populations of European ancestry. Aside from replication of the *TERT* and *TERC* association signals already described, only rs8105767 in the *ZNF208* region was observed to have a nominally significant P value ( $p=0.003$ ) (see online supplementary table S6). However, the direction of association was consistent with prior studies for all nine SNPs ( $P=0.002$ , based on a binomial distribution).<sup>21</sup>

The SNP-based  $h^2$  estimate for TL using data on all subjects (including close relatives) was 27.5% ( $P=1.2 \times 10^{-10}$ ) (table 3). We were unable to generate a non-zero  $h^2$  estimate using only ‘unrelated’ individuals, likely due to the small sample size at that threshold ( $n=1524$ , with no pairs having a coefficient of relationship  $<0.05$ ). When we excluded first cousins and closer related individuals, the  $h^2$  estimate was 6%, although it was not significant ( $P=0.296$ ). In analyses restricting to pairs with  $r \leq 0.25$  (ie, excluding half sibs, aunts/uncles and closer relatives), the  $h^2$  estimate was 26% ( $P=0.0006$ ). The SNP-based narrow-sense  $h^2$  estimate using the method described by Zaitlen *et al*<sup>43</sup> was 44.2% ( $P=2.7 \times 10^{-14}$ ). This estimate captures the effects of rare variants but is prone to upward bias due to the effects of shared environment.

## DISCUSSION

In this paper, we report the first GWAS of TL conducted in a South Asian population. Prior GWAS of TL conducted primarily among individuals of European ancestry have reported significant associations at 10 loci: *ACYP2* (2p16.2),<sup>21</sup> *TERC* (3q26.2),<sup>20 21 26–29</sup> *NAF1* (4q32.2),<sup>21</sup> *TERT* (5p15.33),<sup>21</sup> *OBFC1* (10q24.3),<sup>23</sup> *DCAF4* (14q24.2),<sup>24</sup> *CTC1* (17p13.1),<sup>25</sup> *ZNF208* (19p12),<sup>21</sup> *ZNF676* (19p12)<sup>25</sup> and *RTEL1* (20q13.3).<sup>21</sup> Our study, conducted in a homogenous Bangladeshi population,

replicated two TL-associated SNPs (in the *TERC* and *TERT* regions) and identified a novel association signal in the *RTEL1* region, which is independent of the previously identified association signal in this region. Among the seven remaining previously reported loci, all showed associations that were directionally consistent with prior studies, but only one showed a nominally significant ( $P < 0.05$ ) association (rs8105767 (*ZNF208*)  $P=0.003$ ). We observed five suggestive association signals that warrant follow-up in future studies, especially rs1151814 near the poly(ADP-ribose)polymerase 1 gene (*PARP1*), which is preferentially recruited to damaged telomeres by interacting with telomere repeat binding factor 2 (*TRF2*) and modifying the binding of *TRF2* to telomeric DNA.<sup>44</sup>

While we did not replicate the association for the previously reported SNP in the *RTEL1* region, the novel association signal we observe is best represented by rs2297439, which lies in intron 12 of *RTEL1*. In contrast, the previously reported signal near *RTEL1* is best represented by rs755017, which is located ~140 kb downstream of the *RTEL1* start site in exon 2 of *ZBTB46*. In both the BEB and EUR 1000 Genomes populations, the SNPs in rs755017 peak are not in LD with the SNPs in the rs2297439 peak (see online supplementary figure S6). Additionally, the MAF of rs2297439 in our Bangladeshi population (MAF=0.25) is substantially larger than in populations of European ancestry, ranging from 0.24 to 0.39 in the five South Asian 1000 Genomes populations (Gujarati Indian from Houston (GIH), Punjabi from Lahore, Pakistan (PJI), Bengali from Bangladesh (BEB), Sri Lankan Tamil from the UK (STU), and Indian Telugu from the UK (ITU)) to 0.11 in Southern Europe (Toscani in Italy (TSI) and Iberian Population in Spain (IBS)), 0.07 in Northern Europe (CEU+FIN+GBR) and  $\geq 0.07$  in all other 1000 Genomes populations except Puerto Ricans from Puerto Rico (PUR) (MAF=0.12). This SNP was not reported in the ENGAGE consortium GWA study results. It is possible that we are underpowered to detect

**Table 3** SNP-based heritability analysis with varying coefficients of relationship ( $r$ ) thresholds using genotyped SNPs (~246K)

n	Relationship ( $r$ ) threshold	Heritability estimates (SE)	P	Description of relationships analysed
1524	0.05	0 (0.19)	0.5	Excludes second, third and fourth cousins and closer
2471	0.125	0.06 (0.11)	0.296	Excludes first cousins and closer
3006	0.25	0.26 (0.09)	0.0006	Excludes half sibs, aunt/uncle and closer
3168	0.375	0.19 (0.08)	0.007	Excludes three-quarters sibs or sibling-cousins and closer
5069	None	0.28 (0.05)	$1.2 \times 10^{-10}$	All relationships included
5069	None	0.44 (0.06)*	$2.7 \times 10^{-14}$	All relationships included but distant relatives ( $r < 0.05$ ) set to zero
Jointly estimated*				
5069	None	$1 \times 10^{-6}$ (0.06)	1	All relationships included
5069	None	0.44 (0.08)	$1.4 \times 10^{-7}$	All relationships included but distant relatives ( $r < 0.05$ ) set to zero

\*Heritability estimate calculated using method described by Zaitlen *et al*.<sup>43</sup>

an association for the SNP in the *RTEL1* region reported by Codd *et al.*<sup>21</sup> (rs755017), as the Codd *et al.* study reported a P value of  $6.7 \times 10^{-9}$  using a sample size of 37 684.<sup>22</sup>

Our lead *RTEL1* SNP (rs2297439) resides in intron 12 of *RTEL1*. It also lies within a cis-eQTL (expression quantitative trait loci) affecting expression of *STMN3* in transformed fibroblast cells ( $\beta = -0.68$ ,  $P = 2.8 \times 10^{-9}$ ) (GTEx Analysis Release V6p) and is in LD with the lead SNP for this eQTL ( $r^2 = 0.87$ ). However, this SNP is not reported as a cis-eQTL in any other tissue type. Among the 10 SNPs in high LD ( $r^2 > 0.8$ ) with rs2297439, six reside in regions associated with enhancer histone marks across multiple tissue types (rs77552606, rs75031349, rs78430218, rs16983884, rs76636901 and rs143190905), and three of these variants (rs78430218, rs75031349 and rs76636901) are also located in regions associated with promoter histone marks in multiple tissues.<sup>45</sup> Our lead SNP (rs2297439) is not in strong LD with the nearby variants that have been reported to be associated with risk for glioma (rs6010620,  $r^2 = 0.16$ ) or ulcerative colitis (rs2315008,  $r^2 = 0.10$ ).

The variants we identified in the *TERC* (rs12638862) and *TERT* regions (rs7705526) are well known to be associated with leucocyte TL, and these genes have well-known roles in telomere maintenance. Our lead SNP in *TERC* has not been reported previously but is in LD with several additional previously reported lead SNPs, including rs10936599 ( $r^2 = 0.87$ ) and rs3772190 ( $r^2 = 0.84$ ), but less so with rs16847897 ( $r^2 = 0.27$ ) (based on BEB 1000 Genomesdata). Our lead SNP in *TERT* has been reported previously in a Korean cohort.<sup>22</sup> Previously reported *TERT* SNP rs2736100 is in LD with our lead SNP ( $r^2 = 0.57$ ). Our lead *TERC* and *TERT* SNPs are located in intronic regions of *TERC* and *TERT*, respectively. Our lead *TERC* SNP (rs12638862) lies within cis-eQTLs affecting expression of *RP11-362K14.6* in brain ( $P = 2.0 \times 10^{-7}$ ) and transformed fibroblasts ( $P = 5.9 \times 10^{-5}$ ), *LRRC34* in transformed fibroblast ( $P = 7.8 \times 10^{-6}$ ), nerve ( $P = 2.6 \times 10^{-5}$ ), subcutaneous adipose ( $P = 2.8 \times 10^{-5}$ ) and heart ( $P = 4.8 \times 10^{-5}$ ), *MYNN* in testis ( $P = 1.8 \times 10^{-5}$ ) and *LRR1Q4* in testis ( $P = 6.9 \times 10^{-5}$ ). However, our lead *TERT* SNP (rs7705526) does not appear to be a known cis-eQTL (based on GTEx Analysis Release V6p).

TL has been observed to be highly heritable.<sup>17 46–49</sup> In our study population, which includes some close relatives, our SNP-based  $h^2$  estimate was 27.5% ( $P = 1.2 \times 10^{-10}$ ). The full narrow-sense  $h^2$  estimate was 44% ( $P = 2.74 \times 10^{-14}$ ), using the method described by Zaitlen *et al.*<sup>43</sup> However, both of these estimates are prone to upward bias due to shared environment. While we were unable to estimate the  $h^2$  due to measured SNPs among unrelated individuals (due to small sample size), a recent study of unrelated older adults estimated that 28% of the variance of TL can be explained by common SNPs.<sup>47</sup> Heritability estimates from prior twin studies are as high as 80% and range from 30% to 50% in sibling studies.<sup>46 48 49</sup> A meta-analysis of six family-based studies found a substantial genetic component ( $h^2 = 0.70$ ,  $P = 2.31 \times 10^{-111}$ ).<sup>17</sup> However, family-based estimates could be inflated by shared environment and may capture the effects of ‘trans-generational’ (ie, direct) transmission of telomeres from parent to offspring,<sup>50</sup> not just the  $h^2$  due to inherited variation in non-telomeric sequence.

One limitation of the present study is using different methods to measure TL across cohorts (qPCR vs Luminex based). While both methods performed in our lab show strong correlations with Southern blot (the gold standard for TL measurement),<sup>38 40</sup> the use of multiple methods can introduce heterogeneity into our TL measure. We addressed this limitation by implementing a mixed-effects model to remove experimental variation from both qPCR and Luminex-based measures and standardised the

data to produce similar distributions of TL across all batches and methods. Additionally, while both of these methods are high-throughput and cost-efficient for large epidemiological studies, the values obtained are relative measures of average TL, as opposed to absolute measures. An additional limitation is our lack of a replication data set from this population; however, since our results focus primarily on regions well-known to affect TL, replication may not be as critical for our findings. A major strength of this study is the uniqueness and homogeneity of our study population, which provides the first genome-wide study of TL in a South Asian population.

In summary, this GWAS of TL in a South Asian population replicated two association signals observed in prior GWAS of European ancestry populations. The effects of variants in the *TERC* and *TERT* regions appear to be consistent across European and South Asian ancestry groups. Ancestry-based differences in allele frequencies allowed us to detect a novel association in the *RTEL1* region, which is distinct from the prior association signal reported in this region. Additional research is needed to enhance our understanding of ancestry-based differences in the genetic architecture of TL.

#### Author affiliations

- <sup>1</sup>Department of Public Health Sciences, University of Chicago, Chicago, Illinois, USA
- <sup>2</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA
- <sup>3</sup>Department of Psychiatry, University of California San Diego, La Jolla, California, USA
- <sup>4</sup>Division of Foodborne, Waterborne, and Environmental Diseases, Center for Disease Control, Atlanta, Georgia, USA
- <sup>5</sup>Division of Epidemiology and Biostatistics, University of Illinois, Chicago, Illinois, USA
- <sup>6</sup>UChicago Research Bangladesh, Dhaka, Bangladesh
- <sup>7</sup>Research and Evaluation Division, BRAC, Dhaka, Bangladesh
- <sup>8</sup>Health Systems and Population Studies Division, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh
- <sup>9</sup>Department of Human Genetics, University of Chicago, Chicago, Illinois, USA
- <sup>10</sup>Comprehensive Cancer Center, University of Chicago, Chicago, Illinois, USA
- <sup>11</sup>Department of Medicine, University of Chicago, Chicago, Illinois, USA

**Acknowledgements** The authors would like to sincerely thank all study participants and research staff that have contributed to this research.

**Contributors** Design of telomere length studies within Heals and Best: BLP; design of Heals and Best studies: HA; generation of genetic and molecular data: MGK, SR, MR, JS and MS; collection of data from study participants: MA, MR-Z, AA, TI, GS, MR and MY; quality control for telomere length measures: CZ; SNP data quality control and imputation: CZ and LT; data analysis and interpretation: BLP and DD; drafting the article: DD; critical revision of the article: all authors.

**Funding** This work was supported by the National Institutes of Health (R01ES020506, U01 HG007601, P42 ES10349, R01 CA107431, R01 CA102484, P30CA014599).

**Competing interests** None declared.

**Ethics approval** The institutional review boards of the University of Chicago and Columbia University, and the ethical committee of the Bangladesh Medical Research Council reviewed and approved these projects.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

#### REFERENCES

- 1 Blackburn EH, Epel ES, Lin J. Human telomere biology: a contributory and interactive factor in aging, disease risks, and protection. *Science* 2015;350:1193–8.
- 2 Shawi M, Autexier C. Telomerase, senescence and ageing. *Mech Ageing Dev* 2008;129:3–10.

- 3 Cong YS, Wright WE, Shay JW. Human telomerase and its regulation. *Microbiol Mol Biol Rev* 2002;66:407–25.
- 4 Huang Y, Liang P, Liu D, Huang J, Songyang Z. Telomere regulation in pluripotent stem cells. *Protein Cell* 2014;5:194–202.
- 5 Kalmbach K, Robinson LG, Wang F, Liu L, Keefe D. Telomere length reprogramming in embryos and stem cells. *Biomed Res Int* 2014;2014:1–7.
- 6 Aubert G, Lansdorp PM. Telomeres and aging. *Physiol Rev* 2008;88:557–79.
- 7 Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. *The Lancet* 2003;361:393–5.
- 8 Haycock PC, Heydon EE, Kaptoge S, et al. Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ* 2014;349:g4227.
- 9 Willeit P, Willeit J, Brandstätter A, Ehrlénbach S, Mayr A, Gasperi A, Weger S, Oberhollenzer F, Reinndl M, Kronenberg F, Kiechl S. Cellular aging reflected by leukocyte telomere length predicts advanced atherosclerosis and cardiovascular disease risk. *Arterioscler Thromb Vasc Biol* 2010;30:1649–56.
- 10 Willeit P, Willeit J, Mayr A, Weger S, Oberhollenzer F, Brandstätter A, Kronenberg F, Kiechl S. Telomere length and risk of incident cancer and cancer mortality. *JAMA* 2010;304:69–75.
- 11 Iles MM, Bishop DT, Taylor JC, Hayward NK, Brossard M, Cust AE, Dunning AM, Lee JE, Moses EK, Akslen LA, Andresen PA, Avril MF, Azizi E, Scarrà GB, Brown KM, Dębniak T, Elder DE, Friedman E, Ghiorzo P, Gillanders EM, Goldstein AM, Gruis NA, Hansson J, Harland M, Helsing P, Hočevar M, Höiom V, Ingvar C, Kanetsky PA, Landi MT, Lang J, Lathrop GM, Lubiński J, Mackie RM, Martin NG, Molven A, Montgomery GW, Novaković S, Olsson H, Puig S, Puig-Butillé JA, Radford-Smith GL, Ransderson-Moor J, van der Stoep N, van Doorn R, Whittemore DC, MacGregor S, Pooley KA, Ward SV, Mann GJ, Amos CI, Pharoah PD, Demenais F, Law MH, Newton Bishop JA, Barrett JH. AMFS Investigators|BD investigators|QMEGA and QTWIN Investigators|SDH Study Group|GenoMEL Consortium. The effect on melanoma risk of genes previously associated with telomere length. *J Natl Cancer Inst* 2014;106.
- 12 Walsh KM, Codd V, Rice T, Nelson CP, Smirnov IV, McCoy LS, Hansen HM, Elhauge E, Ojha J, Francis SS, Madsen NR, Bracci PM, Pico AR, Molinaro AM, Tihan T, Berger MS, Chang SM, Prados MD, Jenkins RB, Wiemels JL, Samani NJ, Wiencke JK, Wrensch MR. ENGAGE Consortium Telomere Group. Longer genotypically-estimated leukocyte telomere length is associated with increased adult glioma risk. *Oncotarget* 2015;6:42468–77.
- 13 Walsh KM, Whitehead TP, de Smith AJ, Smirnov IV, Park M, Endicott AA, Francis SS, Codd V, Samani NJ, Metayer C, Wiemels JL. ENGAGE consortium telomere group. Common genetic variants associated with telomere length confer risk for neuroblastoma and other childhood cancers. *Carcinogenesis* 2016;37:576–82.
- 14 Zhang C, Doherty JA, Burgess S, Hung RJ, Lindström S, Kraft P, Gong J, Amos CI, Sellers TA, Monteiro AN, Chenevix-Trench G, Bickeböller H, Risch A, Brennan P, Mckay JD, Houlston RS, Landi MT, Timofeeva MN, Wang Y, Heinrich J, Kote-Jarai Z, Eeles RA, Muir K, Wiklund F, Grönberg H, Berndt SI, Chanock SJ, Schumacher F, Haiman CA, Henderson BE, Amin Al Olama A, Andrlis IL, Hopper JL, Chang-Claude J, John EM, Malone KE, Gammon MD, Ursin G, Whittemore AS, Hunter DJ, Gruber SB, Knight JA, Hou L, Le Marchand L, Newcomb PA, Hudson TJ, Chan AT, Li L, Woods MO, Ahsan H, Pierce BL. GECCO and GAME-ON Network: CORECT, DRIVE, ELLIPSE, FOCI, and TRICL. Genetic determinants of telomere length and risk of common cancers: a Mendelian randomization study. *Hum Mol Genet* 2015;24:5356–66.
- 15 Ojha J, Codd V, Nelson CP, Samani NJ, Smirnov IV, Madsen NR, Hansen HM, de Smith AJ, Bracci PM, Wiencke JK, Wrensch MR, Wiemels JL, Walsh KM. ENGAGE consortium telomere group. Genetic variation associated with longer telomere length increases risk of chronic lymphocytic leukemia. *Cancer Epidemiol Biomarkers Prev* 2016;25:1043–9.
- 16 Bischoff C, Graakjaer J, Petersen HC, Jeune B, Bohr VA, Koelvræ S, Christensen K. Telomere length among the elderly and oldest-old. *Twin Res Hum Genet* 2005;8:425–32.
- 17 Broer L, Codd V, Nyholt DR, Deelen J, Mangino M, Willemsen G, Albrecht E, Amin N, Beekman M, de Geus EJ, Henders A, Nelson CP, Steves CJ, Wright MJ, de Craen AJ, Isaacs A, Matthews M, Moayyeri A, Montgomery GW, Oostra BA, Vink JM, Spector TD, Slagboom PE, Martin NG, Samani NJ, van Duijn CM, Boomsma DI. Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet* 2013;21:1163–8.
- 18 Hjelmborg JB, Dalgård C, Möller S, Steenstrup T, Kimura M, Christensen K, Kyvik KO, Aviv A. The heritability of leukocyte telomere length dynamics. *J Med Genet* 2015;52:297–302.
- 19 Honig LS, Kang MS, Cheng R, Eckfeldt JH, Thyagarajan B, Leiendecker-Foster C, Province MA, Sanders JL, Perls T, Christensen K, Lee JH, Mayeux R, Schupf N. Heritability of telomere length in a study of long-lived families. *Neurobiol Aging* 2015;36:2785–90.
- 20 Codd V, Mangino M, van der Harst P, Braund PS, Kaiser M, Beveridge AJ, Rafelt S, Moore J, Nelson C, Soranzo N, Zhai G, Valdes AM, Blackburn H, Mateo Leach I, de Boer RA, Kimura M, Aviv A, Goodall AH, Ouwehand W, van Veldhuisen DJ, van Gilst WH, Navis G, Burton PR, Tobin MD, Hall AS, Thompson JR, Spector T, Samani NJ. Wellcome trust case control consortium. Common variants near TERC are associated with mean telomere length. *Nat Genet* 2010;42:197–9.
- 21 Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, Hottenga JJ, Fischer K, Esko T, Surakka I, Broer L, Nyholt DR, Mateo Leach I, Salo P, Hägg S, Matthews MK, Palmén J, Norata GD, O'Reilly PF, Saleheen D, Amin N, Balmforth AJ, Beekman M, de Boer RA, Böhringer S, Braund PS, Burton PR, de Craen AJ, Denniff M, Dong Y, Douroudis K, Dubinina E, Eriksson JG, Garlaschelli K, Guo D, Hartikainen AL, Henders AK, Houwing-Duistermaat JJ, Kananen L, Karssen LC, Kettunen J, Klopp N, Lagou V, van Leeuwen EM, Madden PA, Mägi R, Magnusson PK, Männistö S, McCarthy MI, Medland SE, Mihailov E, Montgomery GW, Oostra BA, Palotie A, Peters A, Pollard H, Pouta A, Prokopenko I, Ripatti S, Salomaa V, Suchiman HE, Valdes AM, Verweij N, Viñuela A, Wang X, Wichmann HE, Widen E, Willemsen G, Wright MJ, Xia K, Xiao X, van Veldhuisen DJ, Catapano AL, Tobin MD, Hall AS, Blakemore AI, van Gilst WH, Zhu H, Erdmann J, Reilly MP, Kathiresan S, Schunkert H, Talmud PJ, Pedersen NL, Perola M, Ouwehand W, Kaprio J, Martin NG, van Duijn CM, Hovatta I, Gieger C, Metspalu A, Boomsma DI, Jarvelin MR, Slagboom PE, Thompson JR, Spector TD, van der Harst P, Samani NJ. CARDIoGRAM consortium. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* 2013;45:422–7.
- 22 Do SK, Yoo SS, Choi YY, Choi JE, Jeon HS, Lee WK, Lee SY, Lee J, Cha SI, Kim CH, Park JY. Replication of the results of genome-wide and candidate gene association studies on telomere length in a Korean population. *Korean J Intern Med* 2015;30:719–26.
- 23 Levy D, Neuhausen SL, Hunt SC, Kimura M, Hwang SJ, Chen W, Bis JC, Fitzpatrick AL, Smith E, Johnson AD, Gardner JP, Srinivasan SR, Schork N, Rotter JJ, Herbig U, Psaty BM, Sastrasiin M, Murray SS, Vasani RS, Province MA, Glazer NL, Lu X, Cao X, Kronmal R, Mangino M, Soranzo N, Spector TD, Berenson GS, Aviv A. Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc Natl Acad Sci U S A* 2010;107:9293–8.
- 24 Mangino M, Christiansen L, Stone R, Hunt SC, Horvath K, Eisenberg DT, Kimura M, Petersen I, Kark JD, Herbig U, Reiner AP, Benetos A, Codd V, Nyholt DR, Sinnreich R, Christensen K, Nassar H, Hwang SJ, Levy D, Bataille V, Fitzpatrick AL, Chen W, Berenson GS, Samani NJ, Martin NG, Tishkoff S, Schork NJ, Kyvik KO, Dalgård C, Spector TD, Aviv A. DCAF4, a novel gene associated with leukocyte telomere length. *J Med Genet* 2015;52:157–62.
- 25 Mangino M, Hwang SJ, Spector TD, Hunt SC, Kimura M, Fitzpatrick AL, Christiansen L, Petersen I, Elbers CC, Harris T, Chen W, Srinivasan SR, Kark JD, Benetos A, El Shami S, Visvikis-Siest S, Christensen K, Berenson GS, Valdes AM, Viñuela A, Garcia M, Arnett DK, Broeckel U, Province MA, Pankow JS, Kammerer C, Liu Y, Nalls M, Tishkoff S, Thomas F, Ziv E, Psaty BM, Bis JC, Rotter JJ, Taylor KD, Smith E, Schork NJ, Levy D, Aviv A. Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Hum Mol Genet* 2012;21:5385–94.
- 26 Pooley KA, Bojesen SE, Weischer M, Nielsen SF, Thompson D, Amin Al Olama A, Michailidou K, Tyrer JP, Benlloch S, Brown J, Audley T, Luben R, Khaw KT, Neal DE, Haimdou F, Donovan JL, Kote-Jarai Z, Baynes C, Shah M, Bolla MK, Wang Q, Dennis J, Dicks E, Yang R, Rudolph A, Schildkraut J, Chang-Claude J, Burwinkel B, Chenevix-Trench G, Pharoah PD, Berchuck A, Eeles RA, Easton DF, Dunning AM, Nordestgaard BG. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. *Hum Mol Genet* 2013;22:5056–64.
- 27 Prescott J, Kraft P, Chasman DI, Savage SA, Mirabello L, Berndt SI, Weissfeld JL, Han J, Hayes RB, Chanock SJ, Hunter DJ, De Vivo I. Genome-wide association study of relative telomere length. *PLoS One* 2011;6:e19635.
- 28 Shen Q, Zhang Z, Yu L, Cao L, Zhou D, Kan M, Li B, Zhang D, He L, Liu Y. Common variants near TERC are associated with leukocyte telomere length in the Chinese Han population. *Eur J Hum Genet* 2011;19:721–3.
- 29 Soerensen M, Thinggaard M, Nygaard M, Dato S, Tan Q, Hjelmborg J, Andersen-Ranberg K, Stevnsner T, Bohr VA, Kimura M, Aviv A, Christensen K, Christiansen L. Genetic variation in TERT and TERC and human leukocyte telomere length and longevity: a cross-sectional and longitudinal analysis. *Aging Cell* 2012;11:223–7.
- 30 Hansen ME, Hunt SC, Stone RC, Horvath K, Herbig U, Ranciaro A, Hirbo J, Beggs W, Reiner AP, Wilson JG, Kimura M, De Vivo I, Chen MM, Kark JD, Levy D, Nyambo T, Tishkoff SA, Aviv A. Shorter telomere length in Europeans than in Africans due to polygenic adaptation. *Hum Mol Genet* 2016;25:2324–30.
- 31 Hunt SC, Chen W, Gardner JP, Kimura M, Srinivasan SR, Eckfeldt JH, Berenson GS, Aviv A. Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. *Aging Cell* 2008;7:451–8.
- 32 Ahsan H, Chen Y, Parvez F, Argos M, Hussain AI, Momotaj H, Levy D, van Geen A, Howe G, Graziano J. Health Effects of Arsenic Longitudinal Study (HEALS): description of a multidisciplinary epidemiologic investigation. *J Expo Sci Environ Epidemiol* 2006;16:191–205.
- 33 Argos M, Rahman M, Parvez F, Dignam J, Islam T, Quasem I, K Hore S, T Haider A, Hossain Z, I Patwary T, Rakibuz-Zaman M, Sarwar G, La Porte P, Harjes J, Anton K, Kibriya MG, Jasmine F, Khan R, Kamal M, Shea CR, Yunus M, Baron JA, Ahsan H. Baseline comorbidities in a skin cancer prevention trial in Bangladesh. *Eur J Clin Invest* 2013;43:579–88.

- 34 Pierce BL, Kibriya MG, Tong L, *et al.* Genome-wide association study identifies chromosome 10q24.32 variants associated with arsenic metabolism and toxicity phenotypes in Bangladesh. *PLoS Genet* 2012;8:e1002522.
- 35 Pierce BL, Tong L, Argos M, Gao J, Farzana J, Roy S, Paul-Brutus R, Rahaman R, Rakibuz-Zaman M, Parvez F, Ahmed A, Quasem I, Hore SK, Alam S, Islam T, Harjes J, Sarwar G, Slavkovich V, Gamble MV, Chen Y, Yunus M, Rahman M, Baron JA, Graziano JH, Ahsan H. Arsenic metabolism efficiency has a causal role in arsenic toxicity: Mendelian randomization and gene-environment interaction. *Int J Epidemiol* 2013;42:1862–72.
- 36 Ehrlenbach S, Willeit P, Kiechl S, Willeit J, Reindl M, Schanda K, Kronenberg F, Brandstätter A. Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. *Int J Epidemiol* 2009;38:1725–34.
- 37 Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res* 2002;30:e47.
- 38 Kibriya MG, Jasmine F, Roy S, Ahsan H, Pierce B. Measurement of telomere length: a new assay using QuantiGene chemistry on a Luminex platform. *Cancer Epidemiol Biomarkers Prev* 2014;23:2667–72.
- 39 Kibriya MG, Jasmine F, Roy S, Ahsan H, Pierce BL. Genetic variation associated with longer telomere length increases risk of chronic lymphocytic Leukemia. *PLoS One* 2016;11:e0155548.
- 40 Pierce BL, Jasmine F, Roy S, Zhang C, Aviv A, Hunt SC, Ahsan H, Kibriya MG. Telomere length measurement by a novel Luminex-based assay: a blinded comparison to Southern blot. *Int J Mol Epidemiol Genet* 2016;7:18–23.
- 41 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76–82.
- 42 Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* 2014;46:100–6.
- 43 Zaitlen N, Kraft P, Patterson N, *et al.* Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet* 2013;9:e1003520.
- 44 Gomez M, Wu J, Schreiber V, Dunlap J, Dantzer F, Wang Y, Liu Y. PARP1 Is a TRF2-associated poly(ADP-ribose)polymerase and protects eroded telomeres. *Mol Biol Cell* 2006;17:1686–96.
- 45 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–34.
- 46 Andrew T, Aviv A, Falchi M, Surdulescu GL, Gardner JP, Lu X, Kimura M, Kato BS, Valdes AM, Spector TD. Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. *Am J Hum Genet* 2006;78:480–6.
- 47 Faul JD, Mitchell CM, Smith JA, Zhao W. Estimating telomere length heritability in an unrelated sample of adults: is heritability of telomere length modified by life course socioeconomic status? *Biodemography Soc Biol* 2016;62:73–86.
- 48 Slagboom PE, Droog S, Boomsma DI. Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* 1994;55:876–82.
- 49 Vasa-Nicotera M, Brouillette S, Mangino M, Thompson JR, Braund P, Clemitson JR, Mason A, Bodycote CL, Raleigh SM, Louis E, Samani NJ. Mapping of a major locus that determines telomere length in humans. *Am J Hum Genet* 2005;76:147–51.
- 50 De Meyer T, Vandepitte K, Denil S, De Buyzere ML, Rietzschel ER, Bekaert S. A non-genetic, epigenetic-like mechanism of telomere length inheritance? *Eur J Hum Genet* 2014;22:10–11.