nature genetics

Identification of seven loci affecting mean telomere length and their association with disease

Interindividual variation in mean leukocyte telomere length (LTL) is associated with cancer and several age-associated diseases. We report here a genome-wide meta-analysis of 37,684 individuals with replication of selected variants in an additional 10,739 individuals. We identified seven loci, including five new loci, associated with mean LTL ($P < 5 \times 10^{-8}$). Five of the loci contain candidate genes (*TERC*, *TERT*, *NAF1*, *OBFC1* and *RTEL1*) that are known to be involved in telomere biology. Lead SNPs at two loci (*TERC* and *TERT*) associate with several cancers and other diseases, including idiopathic pulmonary fibrosis. Moreover, a genetic risk score analysis combining lead variants at all 7 loci in 22,233 coronary artery disease cases and 64,762 controls showed an association of the alleles associated with shorter LTL with increased risk of coronary artery disease (21% (95% confidence interval, 5–35%) per standard deviation in LTL, $P = 0.014$). Our findings support a causal role of telomere-length variation in some age-related diseases.

Telomeres are the protein-bound DNA repeat structures at the ends of chromosomes that are important in maintaining genomic stability[1]. They are critical in regulating cellular replicative capacity[2]. During somatic-cell replication, telomere length progressively shortens because of the inability of DNA polymerase to fully replicate the 3′ end of the DNA strand. Once a critically short telomere length is reached, the cell is triggered to enter replicative senescence, which subsequently leads to cell death[1,2]. Conversely, in germ cells and other stem cells that require renewal, telomere length is maintained by the enzyme telomerase, a ribonucleoprotein that contains the RNA template *TERC* and a reverse transcriptase TERT[3]. Both longer and shorter telomere length are associated with increased risk of certain cancers[4,5], and reactivation of telomerase, which bypasses cellular senescence, is a common requirement for oncogenic progression[6]. Therefore, telomere length is an important determinant of telomere function.

Mean telomere length exhibits considerable interindividual variability and has high heritability with estimates varying between 44% and 80% (refs. 7–9). Most of these studies have measured mean telomere length in blood leukocytes. However, there is evidence that, within an individual, mean LTL and telomere length in other tissues are highly correlated[10,11]. In cross-sectional population studies, mean LTL is longer in women than in men and is inversely associated with age (declining by between 20–40 bp per year)[9,12–14]. Shorter age-adjusted and sex-adjusted mean LTL has been found to be associated with risk of several age-related diseases, including coronary artery disease (CAD)[12–15], and has been advanced as a marker of biological aging[16]. However, the extent to which the association of shorter LTL with age-related disorders is causal in nature remains unclear. Identifying genetic variants that affect telomere length and testing their association with disease could clarify any causal role.

So far, common variants at two loci on chromosome 3q26 (*TERC*)[17–19] and chromosome 10q24.33 (*OBFC1*)[18], which explain <1% of the variance in telomere length, have shown a replicated association with mean LTL in genome-wide association studies (GWAS). To identify other genetic determinants of LTL, we conducted a large-scale GWAS meta-analysis of 37,684 individuals from 15 cohorts, followed by replication of selected variants in an additional 10,739 individuals from 6 more cohorts.

Details of the studies included in the GWAS meta-analysis and in the replication phase are provided in the **Supplementary Note**, and key characteristics are summarized in **Supplementary Table 1**. All subjects were of European descent, the majority of the cohorts were population based and three of the replication cohorts were additional subjects from studies used in the meta-analysis. The genotyping platforms and the imputation method (to HapMap 2 build 36) used by each GWAS cohort are summarized in **Supplementary Table 2**. We measured mean LTL in each cohort using a quantitative PCR method and expressed it as a ratio of telomere repeat length to copy number of a single-copy gene (T/S ratio; Online Methods and **Supplementary Note**).

Then we analyzed LTL, adjusted for age, sex and any study-specific covariates, for association with genotype using linear regression in each study and adjusted the results for genomic inflation control factors (**Supplementary Table 2**). We performed an inverse variance–weighted meta-analysis for 2,362,330 SNPs (Online Methods) with correction for the overall genomic inflation control factor ($\lambda = 1.007$; quantile-quantile plot for the meta-analysis is shown in **Supplementary Fig. 1**).

SNPs in seven loci exhibited association with mean LTL at genome-wide significance ($P < 5 \times 10^{-8}$; **Figs. 1**, **2**, **Table 1** and **Supplementary Fig. 2**). The association of the lead SNP on chromosome 2p16.2 (rs11125529) was very close to the threshold for genome-wide significance, and the lead SNP in a locus on 16q23.3 (rs2967374) fell just short of this threshold (**Table 1**). We therefore sought replication of results for these two loci. We confirmed the association of rs11125529

A full list of authors and affiliations appears at the end of the paper.

Received 26 June 2012; accepted 19 December 2012; published online 27 March 2013; doi:10.1038/ng.2528

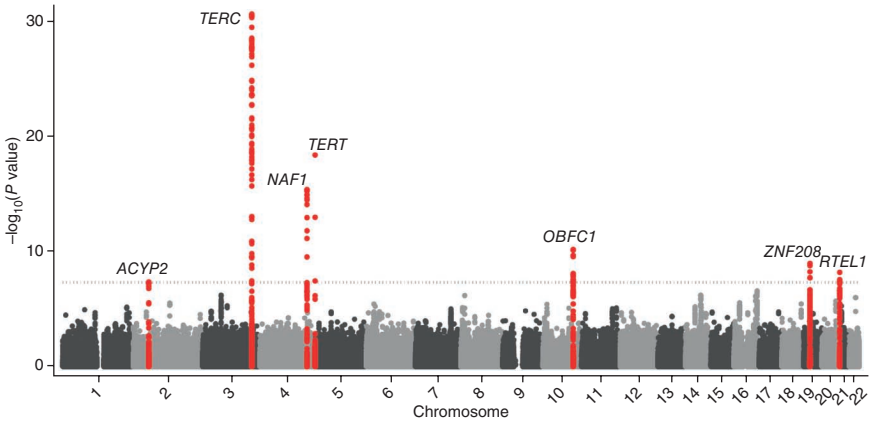

Figure 1 Signal-intensity plot of genotype association with telomere length. Data are displayed as $-\log_{10}(P$ values) against chromosomal location for the 2,362,330 SNPs that were tested. The dotted line represents a genome-wide level of significance at $P = 5 \times 10^{-8}$. Loci that showed an association at this level are plotted in red.

but not of rs2967374 (**Table 1**). The combined *P* value from the GWAS meta-analyses and replication cohorts for rs11125529 was 7.50×10^{-10}. There was no evidence of sex-dependent effects or additional independent signals at any of these loci (Online Methods and **Supplementary Tables 3**, **4**).

Details of key genes in each locus associated with LTL and their location in relation to the lead SNP are provided in **Supplementary Table 5**. The most significantly associated locus we found was the previously reported *TERC* locus on 3q26 (**Figs. 1**, **2** and **Table 1**)[17]. Four additional loci, 5p15.33 (*TERT*), 4q32.2 (*NAF1*, nuclear assembly factor 1), 10q24.33 (*OBFC1*, oligonucleotide/oligosaccharide-binding fold containing 1)[18] and 20q13.3 (*RTEL1*, regulator of telomere elongation helicase 1), harbor genes that encode proteins with known function in telomere biology[3,20–23]. NAF1 protein is required for assembly of H/ACA box small nucleolar RNA, the RNA family to which *TERC* belongs[20]. Thus, the three most significantly associated loci (3q26, 5p15.33 and 4q32.2) harbor genes involved in the formation and activity of telomerase. We therefore examined whether the lead SNPs at these loci as well as the other identified loci associate with leukocyte telomerase activity in available data from 208 individuals. We did not find an association of any of the variants with telomerase activity (**Supplementary Table 6**). However, the study only had 80% power (α of 0.05) to detect a SNP effect that explained 3.7% of the variance in telomerase activity, and therefore smaller effects are likely to have been missed in this exploratory analysis.

We also found a significant association ($P = 6.90 \times 10^{-11}$) at the previously reported *OBFC1* locus[18]. OBFC1 is a component of the telomere-binding CST complex that also contains CTC1 and TEN1 (ref. 21). In yeast, this complex binds to the single-stranded guanine overhang at the telomere and functions to promote telomere replication. RTEL1 is a DNA helicase that has been shown to have important roles in setting telomere length, telomere maintenance and DNA repair in mice[22,23]. However, it should be noted that the

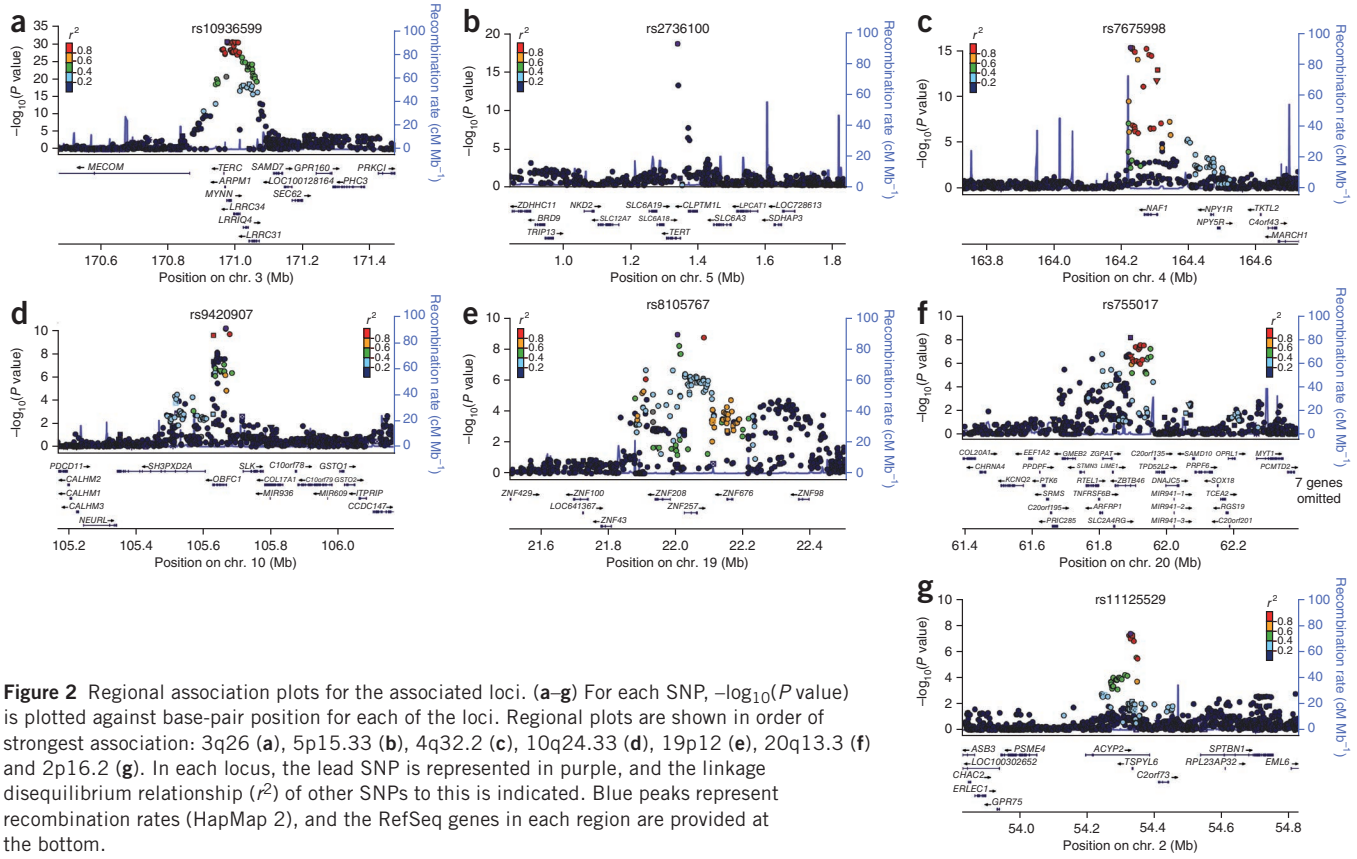

Figure 2 Regional association plots for the associated loci. (**a**–**g**) For each SNP, $-\log_{10}(P$ value) is plotted against base-pair position for each of the loci. Regional plots are shown in order of strongest association: 3q26 (**a**), 5p15.33 (**b**), 4q32.2 (**c**), 10q24.33 (**d**), 19p12 (**e**), 20q13.3 (**f**) and 2p16.2 (**g**). In each locus, the lead SNP is represented in purple, and the linkage disequilibrium relationship (r^2) of other SNPs to this is indicated. Blue peaks represent recombination rates (HapMap 2), and the RefSeq genes in each region are provided at the bottom.

Table 1 Results of telomere length genome-wide association meta-analysis and replication analysis

SNP	Chr.	Position	Gene	*N*	Effect allele	Other allele	Effect allele frequency	β	Standard error	*P* value	Explained variance (%)	Effect on LTL expressed as: equivalent age-related attrition[a]	Effect on LTL expressed as: base pairs[b]
GWAS meta-analysis													
rs10936599	3	170974795	*TERC*	37,669	T	C	0.252	−0.097	0.008	2.54×10^{-31}	0.36	3.91	117.3
rs2736100	5	1339516	*TERT*	25,842	A	C	0.514	−0.078	0.009	4.38×10^{-19}	0.31	3.14	94.2
rs7675998	4	164227270	*NAF1*	34,694	A	G	0.217	−0.074	0.009	4.35×10^{-16}	0.19	2.99	89.7
rs9420907	10	105666455	*OBFC1*	37,653	A	C	0.865	−0.069	0.010	6.90×10^{-11}	0.11	2.76	82.8
rs8105767	19	22007281	*ZNF208*	37,499	A	G	0.709	−0.048	0.008	1.11×10^{-9}	0.09	1.92	57.6
rs755017	20	61892066	*RTEL1*	37,113	A	G	0.869	−0.062	0.011	6.71×10^{-9}	0.09	2.47	74.1
rs11125529	2	54329370	*ACYP2*	37,653	C	A	0.858	−0.056	0.010	4.48×10^{-8}	0.08	2.23	66.9
rs2967374	16	80767362	*MPHOSPH6*	37,437	G	A	0.790	−0.045	0.009	2.70×10^{-7}	NA	NA	NA
Selective replication													
rs11125529	2	54329370	*ACYP2*	10,254	C	A	0.864	−0.053	0.070	4.70×10^{-3}	NA	NA	NA
rs2967374	16	80767362	*MPHOSPH6*	9,063	G	A	0.790	−0.004	0.031	7.80×10^{-1}	NA	NA	NA

N refers to the number of individuals meta-analyzed for each SNP and, for rs11125529 and rs296734, the additional samples used in replication. The sample size for rs2736100 is smaller than for other loci as this SNP is only present on certain genotyping platforms and, because of weak LD structure in the region, cannot be imputed reliably. 'Effect allele' indicates the allele that is associated with shorter telomere length, explaining why all the β estimates are negative. NA, not applicable.
[a]Estimates of the per-allele effect on average age-related telomere attrition in years (based on data in **Supplementary Fig. 3**). [b]Estimates of the per-allele effect on LTL in base pairs calculated from the equivalent age-related attrition in T/S ratio.

lead SNP is 94 kb from *RTEL1*. The remaining two loci (19p21 and 2p16.2) do not harbor obvious candidate genes related to telomere biology. The locus on 19p12 contains a cluster of genes encoding zinc-finger proteins, and the locus on 2p16.2 spans both the *ACYP2* gene, which encodes a muscle-specific acylphosphate, and *TSPYL6*, a gene in intron 3 of *ACYP2* that has homology with nucleosome assembly factor genes. There is evidence that *ACYP2* is linked to stress-induced apoptosis in rat muscle[24].

To gain functional insight into the associated loci, we undertook various bioinformatics analyses (Online Methods). Details of the findings are provided in the **Supplementary Note** and in **Supplementary Table 7**. SNPs in high linkage disequilibrium with the lead SNP were within potential regulatory elements of *TERC*, *NAF1* and *OBFC1*. However, similar SNPs were also present for other genes in some of the loci. These findings emphasize that, although strong candidate genes are located in some of the loci, at this stage we cannot overlook the potential involvement of other genes in each region.

Each of the identified loci explains a relatively small proportion of the total variance in LTL (**Table 1**). To put this in context, we calculated the effect of the lead SNP at each locus in terms of equivalent age-related shortening of LTL based on an estimate of age-related attrition in T/S ratio calculated across all cohorts (**Supplementary Fig. 3**). We saw per-allele effects using this measure equivalent to 1.9–3.9 years of age-related attrition in T/S ratio (**Table 1**). The quantitative PCR method we used here to measure LTL cannot be used to directly calculate the effect on LTL in base pairs. However, many prior studies that have used DNA blotting to measure LTL have shown that mean LTL attrition rate is ~30 bp per year[8,12–14,25]. This suggests that the per-allele effect of the different SNPs on LTL in base pairs ranges from ~57 bp to 117 bp (**Table 1**).

As both shorter and longer mean LTL have been linked to increased risk of various diseases, we searched genetic-association databases for disease associations with the LTL-associated SNPs (**Supplementary Table 8**). The rs10936599 (*TERC*) allele associated with longer LTL associates with increased risk of colorectal cancer[19] and with two autoimmune diseases, multiple sclerosis (longer LTL allele) and celiac disease (shorter LTL allele). The lead SNP for the 5p15.33 (*TERT*) locus is associated with different cancer types (both shorter and longer LTL alleles) and with increased risk of idiopathic pulmonary fibrosis (shorter LTL allele), a disease that has previously been shown to be associated with shorter LTL[26].

One of the most widely reported associations for LTL to date has been that between shorter mean LTL and CAD[12–14,25]. Because LTL is also affected by other risk factors for CAD such as oxidative stress[27–29], it has been unclear whether the association of shorter LTL with CAD is primary or secondary. To investigate whether the association could be causal, we examined the association of both individual lead SNPs and a genetic risk score based on a combination of all 7 SNPs (adjusted for their effect size) with CAD in the CARDIoGRAM GWAS meta-analysis comprising 22,233 CAD cases and 64,762 controls

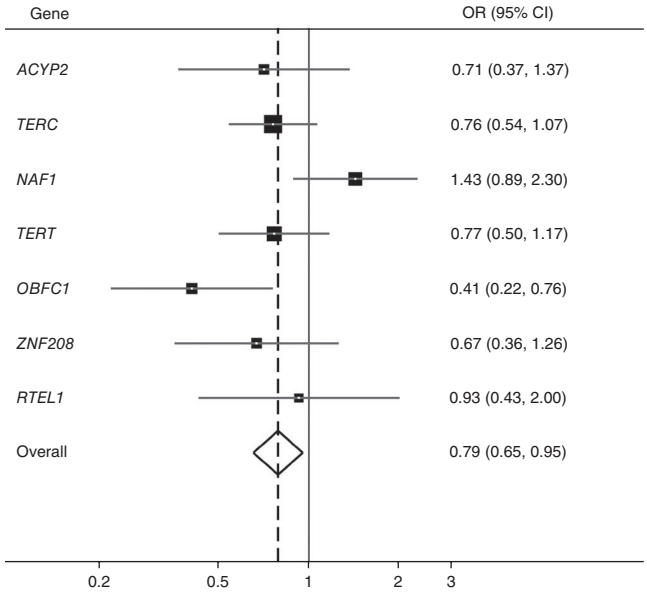

Figure 3 Telomere length variants and risk of CAD. Forest plot showing the effect of telomere length on CAD risk obtained for each SNP using a risk score analysis[31] for each SNP. Effect sizes are plotted with 95% confidence intervals. The overall estimate is from a fixed-effects meta-analysis over all SNPs, where the odds ratio (OR) relates to the change in CAD risk for a s.d. change in telomere length.

of European descent[30], using the approach recently described by the ICBP Consortium[31]. Although the results for individual variants were not significant, 6 of 7 variants showed consistency in direction, and the combined genetic risk score analysis showed a significant association ($P = 0.014$) of the allele associated with shorter LTL with increased risk of CAD (**Fig. 3**). Shorter mean LTL equivalent to one standard deviation in LTL was associated with a 21% (95% confidence interval, 5–35%) higher risk of CAD.

Here we report five new and confirm two previously reported loci that associate with mean LTL in humans. A specific motivation for our study was the observation that variation in LTL is associated with several age-related diseases and the desire to establish whether this link is causal. This is particularly challenging to disentangle because other environmental and lifestyle factors also affect telomere length[29,32–34]. The most persuasive evidence for a causal role comes from *in vitro* and *in vivo* manipulation of telomerase activity, which affects telomere length and has been shown to enhance or reverse senescence and aging-associated phenotypes[35–39]. Here we show that some of the genetic variants associated with LTL are also associated with risk of specific cancers as well as other diseases, some of which have been shown to be previously associated with shorter LTL, suggesting a causal link. An interesting finding was that alleles associated with both shorter and longer telomeres showed associations with specific cancers, suggesting that variation in LTL in either direction may contribute to the development of specific cancers.

As an example of a complex disease that has been shown to be associated with shorter LTL, we examined CAD. Through an analysis of a large GWAS database of CAD[30], we found that, although individually the lead SNPs at each of the telomere length–associated loci were not significantly associated with risk of CAD (probably at least in part reflecting their weak individual effects on LTL and low power), in a combined analysis, alleles associated with shorter LTL were associated with a significantly higher risk of CAD. Because the variants at each of the loci could have other biological effects that could affect their association with CAD through LTL (and possibly explain why the *NAF1* locus may be trending in the opposite direction), some caution is required in the interpretation of this association. Nonetheless, the finding is consistent with that in the prospective WOSCOPS study where, after adjustment for other CAD risk factors, baseline LTL was associated with a 44% higher risk of CAD over the ensuing mean 5.5 years of follow-up in individuals in the tertile with the shortest LTL compared to that with the longest LTL[13]. Our finding here therefore supports a causal association of shorter LTL with CAD, and mechanistic investigation of this relationship is warranted.

In summary, we provide insights into the genetic determination of a structure that is critically involved in genomic stability and cellular function. Our findings suggest that several candidate genes encoding proteins with known function in telomere biology contribute to the LTL associations. The findings provide a framework for a genetic approach to investigating the causal role of telomere length in aging-related diseases.

URLs. R software, http://www.r-project.org/; 1000 Genomes Project, http://www.1000genomes.org/; Genotype-Tissue Expression Project, http://www.genome.gov/gtex/; and UCSC Genome Browser http://genome.ucsc.edu/.

METHODS

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

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

This study was undertaken under the framework of European Union Framework 7 ENGAGE Project (HEALTH-F4-2007-201413). A full list of acknowledgments, including support for each study, is provided in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

V.C. and N.J.S. supervised the overall study. V.C., M.M., T.D.S., P.v.d.H. and N.J.S. designed the study. M.M., T.E., D.R.N., R.A.d.B., G.D.N., D.S., N.A., A.J.B., P.S.B., P.R.B., K.D., M.D., J.G.E., K.G., A.-L.H., A.K.H., L.C. Karssen, J.K., N.K., V.L., I.M.L., E.M.v.L., P.A.M., R.M., P.K.E.M., S.M., M.I.M., S.E.M., E.M., G.W.M., B.A.O., J.P., A. Palotie, A. Peters, Anneli Pouta, I.P., S.R.,V.S., A.M.V., N.V., A.V., H.-E.W., E.W., G.W., M.J.W., K.X., X.X., D.J.v.V., A.L.C., M.D.T., A.S.H., A.I.F.B., P.J.T., N.L.P., M.P., J.D., W.O., J. Kaprio, N.G.M., C.M.v.D., C.G., A.M., D.I.B., M.-R.J., W.H.v.G., P.E.S., T.D.S., P.v.d.H. and N.J.S. contributed to recruitment, study and data management, genotyping and/or imputation of individual studies. V.C., J.L.B., M.K.M., R.A.d.B., J.P., E.D., L.K., H.P., P.T.J. and I.H. measured telomere length. C.P.N., E.A., M.M., J.D., J.L.B., J.J.H., K.F., T.E., I.S., L.B., D.R.N., R.A.d.B., P.S., S.H., G.D.N., P.F.O., I.M.L., S.E.M. and P.v.d.H. undertook association analysis of individual studies; C.P.N., E.A. and J.R.T. carried out the meta-analysis and the additional reported analyses. H.Z., X.W., D.G. and Y.D. provided data on telomerase activity and genotypes. J.E., M.P.R., S.K. and H.S. contributed CAD association data on behalf of CARDIoGRAM. V.C. and N.J.S. prepared the paper together with C.P.N., E.A., M.M. and P.v.d.H. and all authors reviewed the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Blackburn, E.H., Greider, C.W. & Szostak, J.W. Telomeres and telomerase: the path from maize, *Tetrahymena* and yeast to human cancer and aging. *Nat. Med.* **12**, 1133–1138 (2006).
2. Allsopp, R.C. *et al.* Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl. Acad. Sci. USA* **89**, 10114–10118 (1992).
3. Wang, C. & Meier, U.T. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.* **23**, 1857–1867 (2004).
4. Ma, H. *et al.* Shortened telomere length is associated with increased risk of cancer: a meta-analysis. *PLoS ONE* **6**, e20466 (2011).
5. Wentzensen, I.M., Mirabello, L., Pfeiffer, R.M. & Savage, S.A. The association of telomere length and cancer; a meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* **20**, 1238–1250 (2011).
6. Chang, S., Khoo, C.M., Naylor, M.L., Maser, R.S. & DePinho, R.A. Telomere-based crisis: functional differences between telomerase activation and ALT in tumour progression. *Genes Dev.* **17**, 88–100 (2002).
7. Slagboom, P.E., Droog, S. & Boomsma, D.I. Genetic determination of telomere size in humans: a twin study of three age groups. *Am. J. Hum. Genet.* **55**, 876–882 (1994).
8. Njajou, O.T. *et al.* Telomere length is paternally inherited and is associated with parental lifespan. *Proc. Natl. Acad. Sci. USA* **104**, 12135–12139 (2007).
9. Vasa Nicotera, M. *et al.* Mapping of a major locus that determines telomere length in humans. *Am. J. Hum. Genet.* **76**, 147–151 (2005).
10. Wilson, W.R. *et al.* Blood leukocyte telomere DNA content predicts vascular telomere DNA content in humans with and without vascular disease. *Eur. Heart J.* **29**, 2689–2694 (2008).
11. Okuda, K. *et al.* Telomere length in the newborn. *Pediatr. Res.* **52**, 377–381 (2002).
12. Brouilette, S., Singh, R.K., Thompson, J.R., Goodall, A.H. & Samani, N.J. White cell telomere length and risk of premature myocardial infarction. *Arterioscler. Thromb. Vasc. Biol.* **23**, 842–846 (2003).
13. Brouilette, S. *et al.* Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. *Heart* **94**, 422–425 (2008).
14. Fitzpatrick, A.L. *et al.* Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am. J. Epidemiol.* **165**, 14–21 (2007).
15. Benetos, A. *et al.* Short telomeres are associated with increased carotid atherosclerosis in hypertensive subjects. *Hypertension* **43**, 182–185 (2004).
16. Samani, N.J. & van der Harst, P. Biological ageing and cardiovascular disease. *Heart* **94**, 537–539 (2008).
17. Codd, V. *et al.* Common variants near *TERC* are associated with mean telomere length. *Nat. Genet.* **42**, 197–199 (2010).
18. Levy, D. *et al.* Genome-wide association identifies *OBFC1* as a locus involved in human leukocyte telomere biology. *Proc. Natl. Acad. Sci. USA* **107**, 9293–9298 (2010).
19. Jones, A.M. *et al.* Terc polymorphisms are associated both with susceptibility to colorectal cancer and with longer telomeres. *Gut* **61**, 248–254 (2012).

20. Egan, E.D. & Collins, K. An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. *Mol. Cell. Biol.* **32**, 2428–2439 (2012).
21. Miyake, Y. *et al.* RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. *Mol. Cell* **36**, 193–206 (2009).
22. Ding, H. *et al.* Regulation of murine telomere length by Rtel1: an essential gene encoding a helicase-like protein. *Cell* **117**, 873–886 (2004).
23. Barber, L.J. *et al.* RTEL1 maintains genomic stability by suppressing homologous recombination. *Cell* **135**, 261–271 (2008).
24. Kim, J.W., Kwon, O.Y. & Kim, M.H. Differentially expressed genes and morphological changes during lengthened immobilization in rat soleus muscle. *Differentiation* **75**, 147–157 (2007).
25. Farzaneh-Far, R. *et al.* Telomere length trajectory and its determinants in persons with coronary artery disease: longitudinal findings from the heart and soul study. *PLoS ONE* **5**, e8612 (2010).
26. Alder, J.K. *et al.* Short telomeres are a risk factor for idiopathic pulmonary fibrosis. *Proc. Natl. Acad. Sci. USA* **105**, 13051–13056 (2008).
27. Richter, T. & von Zglinicki, T. Continuous correlation between oxidative stress and telomere length shortening in fibroblasts. *Exp. Gerontol.* **41**, 1039–1042 (2007).
28. Valdes, A.M. *et al.* Obesity, cigarette smoking, and telomere length in women. *Lancet* **366**, 662–664 (2005).
29. Bekaert, S. *et al.* Telomere length and cardiovascular risk factors in a middle-aged population free of overt cardiovascular disease. *Aging Cell* **6**, 639–647 (2007).
30. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
31. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
32. Lin, J., Epel, E. & Blackburn, E. Telomeres and lifestyle factors: roles in cellular aging. *Mutat. Res.* **730**, 85–89 (2012).
33. Farzaneh-Far, R. *et al.* Association of marine omega-3 fatty acid levels with telomeric aging in patients with coronary heart disease. *J. Am. Med. Assoc.* **303**, 250–257 (2010).
34. Epel, E.S. *et al.* Accelerated telomere shortening in response to life stress. *Proc. Natl. Acad. Sci. USA* **101**, 17312–17315 (2004).
35. Minamino, T. *et al.* Endothelial cell senescence inhuman atherosclerosis: role of telomere in endothelial dysfunction. *Circulation* **105**, 1541–1544 (2002).
36. Oh, H. *et al.* Telomerase reverse transcriptase promotes cardiac muscle cell proliferation, hypertrophy, and survival. *Proc. Natl. Acad. Sci. USA* **98**, 10308–10313 (2001).
37. Wong, K.K. *et al.* Telomere dysfunction and Atm deficiency compromises organ homeostasis and accelerates ageing. *Nature* **421**, 643–648 (2003).
38. Samper, E., Flores, J.M. & Blasco, M.A. Restoration of telomerase activity rescues chromosomal instability and premature aging in Terc$^{-/-}$ mice with short telomeres. *EMBO Rep.* **2**, 800–807 (2001).
39. Jaskelioff, M. *et al.* Telomerase reactivation reverses tissue degeneration in aged telomerase deficient mice. *Nature* **469**, 102–106 (2011).

Veryan Codd[1,2,67], Christopher P Nelson[1,2,67], Eva Albrecht[3,67], Massimo Mangino[4,67], Joris Deelen[5,6], Jessica L Buxton[7], Jouke Jan Hottenga[8], Krista Fischer[9], Tõnu Esko[9], Ida Surakka[10,11], Linda Broer[6,12,13], Dale R Nyholt[14], Irene Mateo Leach[15], Perttu Salo[11], Sara Hägg[16], Mary K Matthews[1], Jutta Palmen[17], Giuseppe D Norata[18–20], Paul F O'Reilly[21,22], Danish Saleheen[23,24], Najaf Amin[12], Anthony J Balmforth[25], Marian Beekman[5,6], Rudolf A de Boer[15], Stefan Böhringer[26], Peter S Braund[1], Paul R Burton[27], Anton J M de Craen[28], Matthew Denniff[1], Yanbin Dong[29], Konstantinos Douroudis[9], Elena Dubinina[1], Johan G Eriksson[11,30–32], Katia Garlaschelli[19], Dehuang Guo[29], Anna-Liisa Hartikainen[33], Anjali K Henders[14], Jeanine J Houwing-Duistermaat[6,26], Laura Kananen[34,35], Lennart C Karssen[12], Johannes Kettunen[10,11], Norman Klopp[36,37], Vasiliki Lagou[38], Elisabeth M van Leeuwen[12], Pamela A Madden[39], Reedik Mägi[9], Patrik K E Magnusson[16], Satu Männistö[11], Mark I McCarthy[38,40,41], Sarah E Medland[14], Evelin Mihailov[9], Grant W Montgomery[14], Ben A Oostra[12], Aarno Palotie[42,43], Annette Peters[36,44,45], Helen Pollard[1], Anneli Pouta[33,46], Inga Prokopenko[38], Samuli Ripatti[10,11,42], Veikko Salomaa[11], H Eka D Suchiman[5], Ana M Valdes[4], Niek Verweij[15], Ana Viñuela[4], Xiaoling Wang[29], H-Erich Wichmann[47–49], Elisabeth Widen[10], Gonneke Willemsen[8], Margaret J Wright[14], Kai Xia[50], Xiangjun Xiao[51], Dirk J van Veldhuisen[15], Alberico L Catapano[18,52], Martin D Tobin[27], Alistair S Hall[25], Alexandra I F Blakemore[7], Wiek H van Gilst[15], Haidong Zhu[29], CARDIoGRAM consortium[53], Jeanette Erdmann[54], Muredach P Reilly[55], Sekar Kathiresan[56–58], Heribert Schunkert[54], Philippa J Talmud[17], Nancy L Pedersen[16], Markus Perola[9–11], Willem Ouwehand[42,59,60], Jaakko Kaprio[10,61,62], Nicholas G Martin[14], Cornelia M van Duijn[6,12,13], Iiris Hovatta[34,35,62], Christian Gieger[3], Andres Metspalu[9], Dorret I Boomsma[8], Marjo-Riitta Jarvelin[21,22,63–65], P Eline Slagboom[5,6], John R Thompson[27], Tim D Spector[4], Pim van der Harst[1,15,66,67] & Nilesh J Samani[1,2]

[1]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. [2]National Institute for Health Research Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK. [3]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [4]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [5]Section of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands. [6]Netherlands Consortium for Healthy Aging, Leiden University Medical Center, Leiden, The Netherlands. [7]Section of Investigative Medicine, Imperial College London, London, UK. [8]Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands. [9]Estonian Genome Center, University of Tartu, Tartu, Estonia. [10]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [11]Public Health Genomics Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [12]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. [13]Centre for Medical Systems Biology, Leiden, The Netherlands. [14]Queensland Institute of Medical Research, Brisbane, Australia. [15]Department of Cardiology, University of Groningen, University Medical Center, Groningen, The Netherlands. [16]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [17]Institute of Cardiovascular Science, Univerisity College London, London, UK. [18]Department of Pharmacological and Biomolecular Sciences, Università degli Studi di Milano, Milan, Italy. [19]Centro Societa Italiana per lo Studio dell'Aterosclerosi, Bassini Hospital, Cinisello B, Italy. [20]The Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University, London, UK. [21]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London, UK. [22]Medical Research Council–Health Protection Agency Centre for Environment and Health, Faculty of Medicine, Imperial College London, London, UK. [23]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [24]Center for Non-Communicable Diseases, Karachi, Pakistan. [25]Division of Epidemiology, Leeds Institute of Genetics, Health and Therapeutics, School of Medicine, University of Leeds, Leeds, UK. [26]Section of Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands. [27]Department of Health Sciences, University of Leicester, Leicester, UK. [28]Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden,

The Netherlands. [29]Georgia Prevention Institute, Georgia Health Sciences University, Augusta, Georgia, USA. [30]University of Helsinki, Department of General Practice and Primary Health Care, Helsinki, Finland. [31]Folkhälsan Research Center, Helsinki, Finland. [32]Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland. [33]Institute of Clinical Medicine/Obstetrics and Gynecology, University of Oulu, Oulu, Finland. [34]Research Programs Unit, Molecular Neurology, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland. [35]Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland. [36]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [37]Hanover Unified Biobank, Hanover Medical School, Hanover, Germany. [38]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. [39]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. [40]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [41]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK. [42]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK. [43]Department of Medical Genetics, University of Helsinki and the Helsinki University Hospital, Helsinki, Finland. [44]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [45]Munich Heart Alliance, Munich, Germany. [46]National Institute for Health and Welfare, Oulu, Finland. [47]Institute of Epidemiology I, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [48]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig Maximilians Universität, Munich, Germany. [49]KlinikumGrosshadern, Munich, Germany. [50]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA. [51]Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [52]Instituto di Ricovero e Cura a Carattere Scientifico Multimedica, Milan, Italy. [53]A full list of members is provided in the **Supplementary Note**. [54]UniversitätzuLübeck, Medizinische Klinik II, Lübeck, Germany. [55]The Cardiovascular Institute, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [56]Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts, USA. [57]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [58]Program in Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA. [59]Department of Hematology, University of Cambridge, Cambridge, UK. [60]National Health Service Blood and Transplant, Cambridge, UK. [61]University of Helsinki, Hjelt Institute, Department of Public Health, Helsinki, Finland. [62]Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland. [63]Institute of Health Sciences, University of Oulu, Oulu, Finland. [64]Biocenter Oulu, University of Oulu, Oulu, Finland. [65]Department of Lifecourse and Services, National Institute for Health and Welfare, Oulu, Finland. [66]Department of Genetics, University of Groningen, University Medical Center, Groningen, The Netherlands. [67]These authors contributed equally to this work. Correspondence should be addressed to N.J.S. (njs@le.ac.uk).

ONLINE METHODS

Subjects. A total of 37,684 individuals from 15 cohorts were used in the GWAS meta-analysis, along with an additional 10,739 individuals from six cohorts for replication of selected variants. All individuals were of European descent. Full details of the discovery and replication cohorts are given in the **Supplementary Note**, and key characteristics are summarized in **Supplementary Table 1**.

Measurements of telomere length and quality control analysis. Mean LTL was measured using a quantitative PCR–based technique[40,41] in all samples. This method expresses telomere length as a ratio (T/S) of telomere repeat length (T) to copy number of a single copy gene (S), in each sample. To standardize across plates, either a calibrator sample or a standard curve was used for quantification. LTL measurements were made in five separate laboratories. Laboratories used are listed for each cohort in **Supplementary Table 1**, and details for the methods used are provided in the **Supplementary Note**. The majority of the samples (67% of the total) were run in a single laboratory with mean inter-run coefficients of variation for LTL measurements in individual cohorts ranging between 2.7% and 3.9%. The remaining samples were run in four other laboratories (**Supplementary Note**). Mean LTL was first assessed for age-related shortening and for an association of longer LTL with female sex in all cohorts, and showed expected associations (**Supplementary Table 1a**,**b**). Ranges in T/S ratios were found to vary between cohorts measured in different laboratories (**Supplementary Table 1**), largely owing to differences in the calibrator or standard DNA used. We therefore standardized LTL in each cohort using a *Z*-transformation approach. The *Z*-transformation was performed separately for males and females for sex-stratified analysis. Effects of age, adjusted for sex, on LTL were estimated in a multiple-regression model on untransformed and *Z*-transformed telomere length in each study separately and combined using a random-effects meta-analysis in STATA (version 11.2, **Supplementary Fig. 3**).

Genotyping, GWAS analysis and study-level quality control. All discovery cohorts had genome-wide genotype information generated on a standard genotyping platform and include imputed genotypes based on HapMapII CEU build 36 as a reference. Detailed information about individual genotyping platforms, imputation methods and analysis software is provided in **Supplementary Table 2**. Within each cohort, SNP associations with LTL were analyzed by linear regression assuming additive effects with adjustment for age and sex as well as study-specific covariates where appropriate, such as adjustments for family and population structure (**Supplementary Table 2**). All study-specific files underwent extensive quality control procedures before meta-analysis. All files were checked for completeness and plausible descriptive statistics on all variables partly supported by the gwasqc function in the program R. Allele frequencies were checked for compliance with HapMap. In addition to the study-specific quality control filters, we included SNP results of a study in our meta-analysis only if the SNP imputation quality score was >0.5 and if the minor allele frequency was >1%. Only SNPs that were available in >50% of the total sample size over all studies were analyzed, resulting in a total number of 2,362,330 SNPs in the meta-analysis.

Meta-analyses. Meta-analysis of all individual study associations was conducted using inverse variance weighting in STATA. As a measure for between study heterogeneity I^2 was calculated[42]. For SNPs with $I^2 \leq 40\%$, fixed-effects models were applied; random-effects models were applied for SNPs with $I^2 > 40\%$. Fixed-effects results were verified by an independent analyst using METAL[43]. Before meta-analysis, standard errors of each study were genomic control corrected using study-specific lambda estimates as provided in **Supplementary Table 2**. The overall inflation factor lambda of the meta-analyzed results was 1.007. Results were further corrected for this. SNPs showing association with telomere length with $P < 5 \times 10^{-8}$, which corresponds to a Bonferroni correction of one million independent tests, were considered to be statistically significant[44].

Replication study. Replication was sought for two SNPs reaching borderline significant *P* values in the discovery analysis. Additional subsets of NTR and ECGUT along with the Leiden 85-plus study had LTL measurements performed. LTL measurements were available for GRAPHIC, PLIC cohorts and for additional samples of PREVEND. *De novo* genotyping was performed either using a commercial genotyping service (GRAPHIC, PREVEND, KBioscience) or by Taqman genotyping as described previously[45]. In these studies, the same model was applied as in the discovery studies. Single study results were meta-analyzed using inverse variance–weighted fixed-effects models in STATA.

Sex-stratified analysis. Genome-wide associations were additionally conducted separately in women and men to investigate whether sex-specific signals existed. Furthermore, all top SNPs from the overall discovery GWAS were tested for differences between women and men by means of the normally distributed test statistic $(\beta_w - \beta_m)/\mathrm{sqrt}(se_w^2 + se_m^2)$, where se represents standard error, w women and m men. The results of this analysis are given in **Supplementary Table 3**.

Conditional association analysis. Regional association plots were generated using LocusZoom[46] for each of the loci containing significantly associated SNPs. These were assessed to check that additional SNPs in high linkage disequilibrium (LD) with the lead SNP also showed some degree of association with telomere length. This was confirmed, but it was evident that some regions (5p15.33, 10q24.33 and 20q13.3) contained SNPs in low LD with the lead SNP that also showed association to LTL. To assess whether independent signals existed at these loci, conditional analyses were carried out. In a subset of studies, a multiple regression model was calculated for each locus including both SNPs. Adjustments were made in the same way as in the single SNP models. Individual study results were meta-analyzed using fixed-effects in R and compared to the meta-analysis results of single SNP models within the same subset of studies. Independency was defined as the percentage change in the effect estimate between the single and the multiple SNP model being ≤25%. The data are provided in **Supplementary Table 4**.

Calculations of explained variances. Explained variances were calculated based on the effect estimates (β) and allele frequencies (EAF) of each single SNP by $2 \times \mathrm{EAF}(1-\mathrm{EAF}) \times (\beta^2/\mathrm{var})$ as suggested before[47]. The phenotypic variance (var) is equal to 1 as the analysis was performed using Z-transformed telomere length.

Genetic risk scores. To assess the impact of these variants on risk of CAD, we performed a multiple-SNP risk score analysis as previously described[31]. This method is equivalent to a fixed-effects inverse variance–weighted meta-analysis of the ratio between the two traits. Lookups were performed in CARDIoGRAM[30] (β_1) to obtain the effect sizes for the seven SNPs along with the standard errors for CAD risk. These were then converted to a ratio (β_3) along with its standard error using the estimates from the telomere meta-analysis (β_2). We removed the BHF-FHS and NBS data from this analysis because they were included in the CARDIoGRAM analysis and to avoid the possibility of reverse causation given the nature of the BHF-FHS sample. The single SNP results were then meta-analyzed using fixed-effects with inverse-variance weighting. The pooled estimate can be interpreted as the effect of a standard deviation increase in telomere length on the risk of CAD.

Leukocyte telomerase activity assays. Details of the cohort are provided in the **Supplementary Note**. Peripheral blood mononuclear cells (PBMCs) were freshly isolated from whole blood by Ficoll-Pague Premium (Sigma-Aldrich) gradient centrifugation within 1 h after blood draw. Isolated PBMCs were stored in a cryopreservation medium composed of RPMI-1640, 10% dimethyl sulfoxide and 10% FBS at liquid nitrogen tank until additional processing. Telomerase activity was assayed by the Telo TAGGG Telomerase PCR ELISA kit (Roche Applied Science) (TRAP assay) per the manufacturer's protocol using 2×10^{-5} cells per assay. An extract from 2,000 cells was used for TRAP reactions. Sample telomerase activity was expressed as ratio of telomerase activity value divided by control HK293 telomerase activity value from 1,000 cells. Intra-assay coefficient of variance (CV) was 5.9% and inter-assay CV was 4.8%. Telomerase activity was log-transformed to obtain better approximations of the normal distribution before analysis. Association analyses with genotype were performed using regression and an additive model with adjustment for age, sex and ethnicity. The interaction between SNP and ethnicity was also built in the regression model to test

doi:10.1038/ng.2528

whether the effect of the SNP on telomerase activity is ethnicity-dependent. The power of the study to detect a SNP effect on telomerase activity was computed using the Genetic Power Calculator[48].

Bioinformatics analyses. For all analyses, we tested lead SNPs and SNPs with an $r^2 > 0.7$ to the lead SNP identified through the 1000 Genomes study at each locus. Functional predictions of any identified coding variants were carried out using PolyPhen2 (ref. 49) and SIFT[50]. To assess whether any variants influenced gene expression, we searched two available genome-wide gene expression databases, the monocyte genome-wide gene expression data from the Gutenburg Heart Study[51] and the Genotype-Tissue Expression Project (GTEx) database, which includes liver, brain and lymphoblastoid cell types. To identify regulatory variants, we searched ENCODE data in the UCSC Genome Browser database[52] to examine whether any SNPs were located within promoter, enhancer or insulator regions (Chromatin State Segmentation), methylation sites (predicted CpG islands and methylation status of the CpG site using data from the Methyl 450K Bead array data and Bisufite sequencing), conserved elements, conserved transcription factor binding sites and regions of known transcription factor binding as shown by transcription factor chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq).

40. Cawthon, R.M. Telomere measurement by quantitative PCR. *Nucleic Acids Res.* **30**, e47 (2002).
41. Cawthon, R.M. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res.* **37**, e21 (2009).
42. Higgins, J.P. *et al.* Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
43. Willer, C.J. *et al.* METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
44. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
45. Salpea, K.D. *et al.* Association of telomere length with type 2 diabetes, oxidative stress and UCP2 gene variation. *Atherosclerosis* **209**, 42–50 (2010).
46. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
47. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
48. Purcell, S., Cherny, S.S. & Sham, P.C. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149–150 (2003).
49. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
50. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
51. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
52. Rosenbloom, K.R. *et al.* ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res.* **40**, d912–d917 (2012).