A genome-wide association study identifies a novel locus on chromosome 18q12.2 influencing white cell telomere length

M Mangino,1 J B Richards,1,2 N Soranzo,1,3 G Zhai,1 A Aviv,4 A M Valdes,1 N J Samani,5 P Deloukas,3 T D Spector1

ABSTRACT

Background: Telomere length is a predictor for a number of common age related diseases and is a heritable trait. Methods and results: To identify new loci associated with mean leucocyte telomere length we conducted a genome wide association study of 314 075 single nucleotide polymorphisms (SNPs) and validated the results in a second cohort (in for both cohorts combined = 2790). We identified two novel associated variants (rs2162440, p = 2.6 × 10−6; and rs7235755, p = 5.5 × 10−6) on chromosome 18q12.2 in the same region as the VPS34/PIKC3C gene, which has been directly implicated in the pathway controlling telomere length variation in yeast.

Conclusion: These results provide new insights into the pathways regulating telomere homeostasis in humans.

Telomeres are nucleoprotein structures capping and protecting the ends of chromosomes. Because of the “end replication problem”,1 telomeres shorten with each cell division and leucocyte telomere length has been shown to decrease with age at a rate of 20–40 base pairs per year.2,3 Telomere attrition is enhanced by inflammation and oxidative stress and short telomere length is an independent predictor of age related diseases such as hypertension, myocardial infarction, congestive heart failure, vascular dementia, osteoporosis, osteoarthritis and Alzheimer’s disease.4

There is wide inter-individual variability in telomere length at birth and at subsequent ages. Both twin studies and intra-familial correlation analysis have identified a genetic influence (from 40% to 80%) on telomere length variation.4,5 Genome-wide linkage studies have identified a genetic influence (from 40% to 80%) on telomere length variation.4,5 Genome-wide association (GWA) analysis is a powerful tool for unlocking the genetic basis of complex traits and has recently provided novel insights into the genetic architecture of many common diseases and traits.6,7 We therefore undertook a GWA scan to identify common alleles that may influence telomere length. Our findings indicate that single nucleotide polymorphisms (SNPs) rs2162440 and rs7235755 on chromosome 18q12.2 are associated with short telomere length in two independent datasets of European descent.

METHODS

We conducted a two stages GWA study on 2790 individuals from the UK Adult Twin Register (table 1), in which we evaluated 314 075 SNPs. The design and methodology of the GWA study is described in detail elsewhere.7 In brief, the discovery sample consisted of 1625 women from the St Thomas’ UK Adult Twin Registry,7 a large cohort of twins historically developed to study the heritability and genetics of diseases with a higher prevalence among women. The sample is not enriched for any particular disease or trait and is representative of the British general population.4 The replication cohort included 1165 subjects of both genders (table 1) from the UK Twin Registry who were unrelated to the individuals from the discovery sample.

Leucocyte telomere length (LTL) was derived by using Southern blot analysis in duplicate to measure the mean terminal restriction fragment.10 The coefficient of variation for this measurement was 1.5%. Because all the individuals of the discovery cohort were females, telomere length was only adjusted for age. After adjustment, the trait was normally distributed in the sample.

Genomic DNA was subjected to SNP genotyping via the Infinium assay (Illumina, San Diego, California, USA), using three fully compatible BeadChip microarrays (HumanHap500-Duo,
<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Marker</th>
<th>Position</th>
<th>Allele</th>
<th>Coding/UTR</th>
<th>Flanking 5'</th>
<th>Flanking 3'</th>
<th>GWA</th>
<th>p Values replication</th>
<th>p Values combined</th>
<th>p Values combined Locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p</td>
<td>rs2468470</td>
<td>59056107</td>
<td>G</td>
<td>coding</td>
<td>23</td>
<td>10</td>
<td>0.40</td>
<td>0.101 (0.039)</td>
<td>1.47E–01</td>
<td>0.206 (0.021)</td>
</tr>
<tr>
<td>1p</td>
<td>rs2468470</td>
<td>59056107</td>
<td>A</td>
<td>UTR</td>
<td>23</td>
<td>10</td>
<td>0.40</td>
<td>0.101 (0.039)</td>
<td>1.47E–01</td>
<td>0.206 (0.021)</td>
</tr>
<tr>
<td>18p</td>
<td>rs590903</td>
<td>87903435</td>
<td>A</td>
<td>coding</td>
<td>92</td>
<td>129</td>
<td>0.07</td>
<td>0.063 (0.022)</td>
<td>2.69E–02</td>
<td>0.178 (0.063)</td>
</tr>
<tr>
<td>18p</td>
<td>rs590903</td>
<td>87903435</td>
<td>G</td>
<td>coding</td>
<td>92</td>
<td>129</td>
<td>0.07</td>
<td>0.063 (0.022)</td>
<td>2.69E–02</td>
<td>0.178 (0.063)</td>
</tr>
<tr>
<td>22q</td>
<td>rs4060747</td>
<td>139803368</td>
<td>A</td>
<td>coding</td>
<td>74</td>
<td>59</td>
<td>0.43</td>
<td>0.07 (0.019)</td>
<td>2.17E–01</td>
<td>0.031 (0.010)</td>
</tr>
<tr>
<td>22q</td>
<td>rs4060747</td>
<td>139803368</td>
<td>G</td>
<td>coding</td>
<td>74</td>
<td>59</td>
<td>0.43</td>
<td>0.07 (0.019)</td>
<td>2.17E–01</td>
<td>0.031 (0.010)</td>
</tr>
</tbody>
</table>

* GWA, genome-wide association; SE, standard error.
HumanHap300 and HumanHap550), according to the manufacturer’s protocols.

We excluded 733 SNPs that had a low call rate (<90%), 2704 SNPs that had Hardy–Weinberg p values <10^{-4}, and 725 SNPs with minor allele frequencies <1%. We also removed subjects where genotyping failed for >2% of SNPs. We retained for the analysis 98.7% (514 075) of all available SNPs. Statistical analysis was carried out with MERLIN (version 1.1.2)11 using the score test (%fastAssoc), while accounting for family structure and twin zygosity.12

RESULTS
In the discovery sample (n = 1625) the strongest association was recorded for rs7374458 on chromosome 3 (5.20 x 10^{-6}). We also identified 28 SNPs with a p value of <10^{-4} and 316 SNPs with a p value of <10^{-3}. We visually inspected all the signal intensity plots of these SNPs and excluded the markers that had been miscalled (11.3%).

Since none observed p values reached a genome-wide significance level after correcting for multiple testing, we adopted the conservative approach of selecting for replication only those polymorphisms with a p value <10^{-4} that were >100 Kb from other associated SNPs (p <1.0 x 10^{-2}). Following these criteria, we identified 15 associated loci including a total of 41 SNPs with the p values for the lead SNPs ranging from 5.20 x 10^{-6} to 9.7 x 10^{-4} (table 2).

These 41 selected SNPs were genotyped in the replication cohort (n = 1165) using Sequenom iPLEX (San Diego, California, USA) technology. Because the replication cohort included both males and females, LTL values were adjusted for both gender and age. After adjustment the trait was again normally distributed. To control for multiple testing, we used an SNP spectral decomposition method proposed by Nyholt13 and modified by Li and Ji.14 After spectral decomposition of the linkage disequilibrium (LD) matrices of the 41 analysed SNPs, the corrected threshold of statistical significance in the replication stage was estimated at p <2.1 x 10^{-3} which is a conservative correction for the number of independent SNPs tested in the replication sample. The results of the association analysis are reported in table 2 and show that we were able to replicate the association observed in the GWA sample for two SNPs: rs2162440 and rs7235755, both mapping to a 2.2 Kb region of chromosome 18q12.2.

Since the discovery cohort included only females, we also performed a gender-specific analysis on the replication population in order to test if the genetic variants may be associated with telomere lengths only for females. The result showed that for both SNPs the direction of the trend was consistent between genders in the replication cohort (rs2162440: —100 (44) base pairs (bp) for females and —140 (70) bp for males; rs7235755: —94 (42) bp for females and —138 (71) bp for males) and between females of the two cohorts (rs2162440: —104 (29) bp for female in discovery and —100 (44) bp for females in replication; rs7235755: —104 (28) bp for female in discovery and —94 (42) bp for females in replication). Although borderline (due to small sample size), p values were statistically significant for both SNPs in both genders in the replication cohort (rs2162440: females p = 0.012, males p = 0.046; rs7235755: females p = 0.02, males p = 0.049).

The joint analysis of genotyped data from the two cohorts yielded combined p values of 2.6 x 10^{-6} (rs2162440) and 5.50 x 10^{-6} (rs7235755). Our analysis also indicated that the G alleles of both SNPs were associated with shorter telomeres (—106 (22) bp for rs2162440 and —103 (22) bp for rs7235755), extrapolating to an approximate 5 years of telomere erosion based on estimates of loss with age.

DISCUSSION
Although our results are unlikely to be artefacts because the identified SNPs were replicated in two independent cohorts, we do believe that our power for identifying association was reduced by the known limitations of the measurement technique.15 Therefore, we can only detect common variants. Indeed, it is likely that there are more loci with small genetic effect that we did not detect because of the stringent thresholds for statistical significance employed in this study. This would explain why we did not detect loci such as those previously identified on chromosome 12q12.22 and 14q25.2.

According to NCBI build 36, the associated polymorphisms map to a 48 Kb LD block within a gene desert, between the Bruno-like 4 (BRUNOL4, NM_020180) and VPS34 (also known as PIK3C3, NM_002647) genes. The identified SNPs (or another variant present in the LD block) might be influencing the expression of either transcript through long range control, as has been demonstrated for other genes.16 This hypothesis is supported by the observation that the associated 48 Kb LD block lies in a highly conserved genomic segment. The two associated variants map ~70 Kb away from BRUNOL4 and 4.3 Mb away from VPS34. BRUNOL4 is a member of the CELF/ Bruno-like family, which encodes proteins bearing highly conserved RNA recognition motif. RNA binding proteins are important elements that control normal cell functions, regulating events such as RNA processing, mRNA transport, stability and translation. VPS34 is a component of the phosphoinositide (PI) 3 kinase family which includes proteins that regulate several aspects of the cell physiology.17 Interestingly, VPS34 yeast orthologue (Vps54) has been directly involved in the pathway which regulates telomere length variation.18

In conclusion, we provide evidence from two independent cohorts for a new locus on chromosome 18q12.2 associated with short telomere length in humans. These data provide new insights into the likely pathways and mechanisms regulating telomere length in humans.

Funding: This study was funded in part by: the Wellcome Trust; NIHR (TDS), NIHR Biomedical Research Centre (grant to Gys’ and St. Thomas’ Hospitals and King’s College London).

Competing interests: None declared.

REFERENCES


A genome-wide association study identifies a novel locus on chromosome 18q12.2 influencing white cell telomere length

M Mangino, J B Richards, N Soranzo, G Zhai, A Aviv, A M Valdes, N J Samani, P Deloukas and T D Spector

*J Med Genet* 2009 46: 451-454 originally published online April 8, 2009
doi: 10.1136/jmg.2008.064956