



## OPEN

SUBJECT AREAS:  
DNA METHYLATION  
IMPRINTING  
GENETICS  
EPIGENOMICSReceived  
7 February 2014Accepted  
17 April 2014Published  
14 May 2014Correspondence and  
requests for materials  
should be addressed to  
J.L.B. (j.buxton@  
imperial.ac.uk)\* These authors  
contributed equally to  
this work.

† Deceased.

# Human leukocyte telomere length is associated with DNA methylation levels in multiple subtelomeric and imprinted loci

Jessica L. Buxton<sup>1\*</sup>, Matthew Suderman<sup>2,3\*</sup>, Jane J. Pappas<sup>3,4</sup>, Nada Borghol<sup>5</sup>, Wendy McArdle<sup>2</sup>, Alexandra I. F. Blakemore<sup>1</sup>, Clyde Hertzman<sup>6†</sup>, Christine Power<sup>7</sup>, Moshe Szyf<sup>3</sup> & Marcus Pembrey<sup>7</sup>

<sup>1</sup>Section of Investigative Medicine, Department of Medicine, Imperial College London, London W12 0NN, UK, <sup>2</sup>University of Bristol, Bristol, UK, <sup>3</sup>McGill University, Montreal, Quebec, Canada, <sup>4</sup>University of Toronto, Toronto, Ontario, Canada, <sup>5</sup>Lebanese International University and Lebanese University, Beirut, Lebanon, <sup>6</sup>University of British Columbia, Vancouver, British Columbia, Canada, <sup>7</sup>UCL Institute of Child Health, London, UK.

In humans, leukocyte telomere length (LTL) is positively correlated with lifespan, and shorter LTL is associated with increased risk of age-related disease. In this study we tested for association between telomere length and methylated cytosine levels. Measurements of mean telomere length and DNA methylation at >450,000 CpG sites were obtained for both blood (N = 24) and EBV-transformed cell-line (N = 36) DNA samples from men aged 44–45 years. We identified 65 gene promoters enriched for CpG sites at which methylation levels are associated with leukocyte telomere length, and 36 gene promoters enriched for CpG sites at which methylation levels are associated with telomere length in DNA from EBV-transformed cell-lines. We observed significant enrichment of positively associated methylated CpG sites in subtelomeric loci (within 4 Mb of the telomere) ( $P < 0.01$ ), and also at loci in imprinted regions ( $P < 0.001$ ). Our results pave the way for further investigations to help elucidate the relationships between telomere length, DNA methylation and gene expression in health and disease.

Telomeres are protective nucleoprotein structures that cap the ends of linear chromosomes. Vertebrate telomeres are composed of variable numbers of a tandem repeat sequence, (TTAGGG)<sub>n</sub>, bound to the shelterin protein complex<sup>1</sup>. Telomere length is maintained by the action of telomerase in some cell types, notably stem and germ cells. However, in most somatic tissues telomeres shorten with each cell division, a process believed to be accelerated by oxidative stress and inflammation<sup>2–5</sup>. Very short or dysfunctional telomeres trigger replicative senescence, a process that may be activated by a single critically short telomere in a cell<sup>6</sup>.

In humans, mean leukocyte telomere length (LTL) is positively correlated with lifespan<sup>7,8</sup>, and has been proposed as a potential biomarker of biological ageing<sup>9–11</sup>. In support of this hypothesis, mean LTL is generally shorter in adult men than in women at any given age, consistent with shorter average male life expectancy in most populations<sup>12</sup>. Furthermore, numerous studies report associations between LTL and age-related diseases including heart disease, type 2 diabetes (T2D) and cancer<sup>13–16</sup>, although the causal direction of these relationships remains largely unknown.

There is substantial inter-individual variation in both newborn telomere length and telomere length shortening rates throughout life<sup>17,18</sup>. Shorter LTL is associated with several known determinants of ill health, including smoking, excessive alcohol consumption, obesity and chronic life stress such as long-term caring responsibilities or prolonged unemployment<sup>19–23</sup>. There is also evidence for association between maltreatment in childhood and shorter adult LTL<sup>24,25</sup>, suggesting that early-life exposures may have long-term effects on cellular ageing rates.

In addition to multiple environmental factors, human LTL is influenced by genetic variation. Recent genome-wide association studies (GWAS) have revealed associations between LTL and several common genetic variants<sup>26–28</sup>. Some of the genes within these associated loci have a direct role in telomere maintenance. These include the *TERT* and *TERC* genes, which encode components of the telomerase enzyme, and *OBFC1* and *CTC1*, which encode components of the human CST (Ctc1, Stn1 and Ten1) complex recently shown to inhibit telomerase<sup>29</sup>. Despite their known roles in telomere biology, the effect sizes of individual variants in these loci are small.

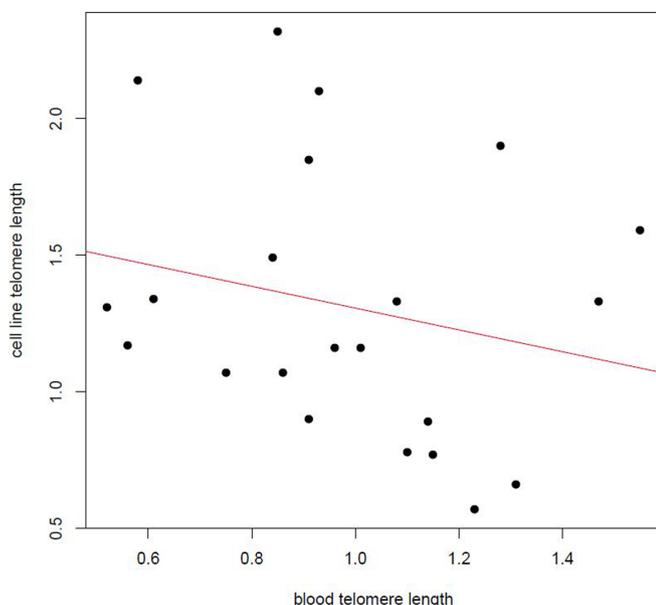


Importantly however, a genetic risk score analysis showed that inheritance of multiple alleles associated with shorter LTL is associated with an increased risk of coronary artery disease, providing preliminary evidence that telomere shortening might play a causal role in this condition<sup>28</sup>.

The aim of the present study was to further investigate the processes involved in telomere shortening by interrogating epigenetic variation in genomic DNA (specifically methylation of cytosine residues), in order to identify novel loci that are either i) affected by telomere shortening or ii) influence telomere length. Since both telomere length and DNA methylation variations are associated with chronological age<sup>30–33</sup>, we compared DNA from individuals all aged 44–45 years when blood samples were taken. We first performed HumanMethylation450 microarray analysis and measured telomere length in DNA from male participants in a longitudinal birth cohort. We next investigated the associations between methylation at CpG sites and telomere length in blood DNA. We also performed a parallel investigation in no-passage EBV transformed cell-lines prepared from the same venepuncture, in order to shed light on the potential mechanisms involved in the maintenance of telomere length in this cell type, and to compare them with any identified in blood cells.

## Results

**Telomere length measurements.** Telomere measurements were obtained for 60 DNA samples from 38 different participants. These comprised 24 blood and 36 cell-line samples (for 22 participants, both blood and cell-line DNA was available). Mean relative telomere length measurements were obtained using the multiplex qPCR method<sup>34</sup>, which provides a “T/S ratio” for each DNA sample. This is a relative measure of the amplification of the telomeric DNA sequence (T) compared to that of a single copy gene (S) in each test sample, normalised using a common reference DNA sample. The whole blood telomere measurements were normally distributed (Mean T/S = 1.00, SD = 0.29, range = 0.52–1.55), whereas those obtained for EBV cell-lines were not normally distributed and show a greater range of values (Mean T/S = 1.26, SD = 0.45, range = 0.57–2.32). We found no evidence of correlation between telomere measurements in whole blood and an EBV cell-line DNA taken from the same individual (Spearman  $\rho = -0.3$ ,  $P = 0.18$ ) (Fig. 1). Six samples showed a decrease in T/S from blood to



**Figure 1** | Scatterplot showing lack of correlation between telomere measurements in whole blood and EBV cell-line DNA taken from the same individual ( $N = 22$ ) (Spearman  $\rho = -0.3$ ,  $P = 0.18$ ).

cell-line; two samples remained same (within 10% of blood T/S value); and 14 samples showed an increase in T/S from blood to cell-line.

**Identification of genes in which DNA methylation is associated with telomere length.** In order to identify genes likely to be affected by telomere length-associated methylation, each CpG represented on the microarray was first assigned the Ensembl gene identifier with the nearest transcription start site (TSS), up to a maximum of 10 Kb away. We then identified gene promoters significantly enriched with CpG sites that are individually associated with telomere length (enrichment false discovery rate  $P < 0.05$  and at least one CpG site with unadjusted association  $P < 0.05$ ; see Methods for more details).

After applying this stringent approach, we found that overall, association statistics were stronger in whole blood DNA than in cell lines, and were mostly positive, i.e. methylation increases as telomere length increases. We identified 364 sites within 47 different gene promoters that were positively associated with telomere length in whole blood DNA. We also found 100 sites that were negatively associated with telomere length, corresponding to 27 different genes (Table 1 and Data S1). Since some gene promoters contained both positively and negatively associated sites, the total number of gene loci for which methylation levels were associated with telomere length in whole blood was 65.

In cell lines, 94 sites were positively associated with telomere length, corresponding to 28 genes, and 31 sites were negatively associated, corresponding to 12 genes (Table 1 and Data S2). Again, as some promoters contained both positively and negatively associated sites, the total number of gene loci for which methylation levels were associated with telomere length in cell-lines was 36.

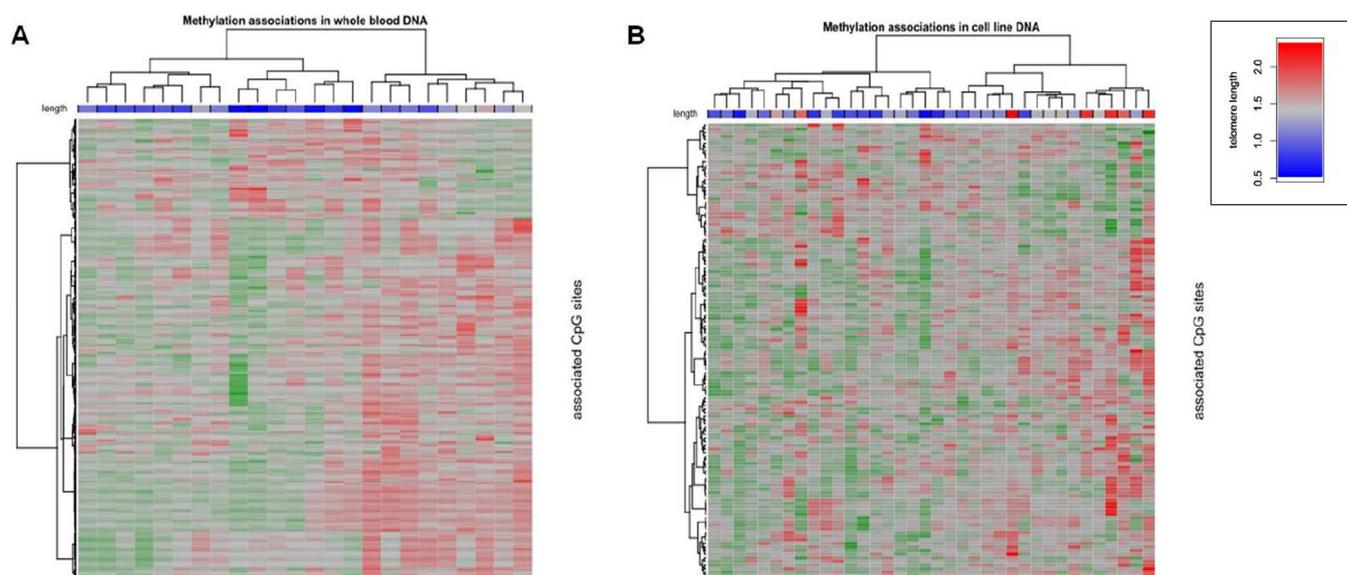
The heatmaps in Fig. 2 illustrate these associations, full details of which are given in Data S1 and S2. Together, these results clearly show that methylation is more strongly associated with telomere length in DNA from whole blood than from cell lines.

**Little in common between cell lines and whole blood.** There are no individual CpG sites associated with telomere length in both whole blood and cell line DNA, although there are three genes with transcription start sites near sites associated in both cell types: *C15orf26*, *RN5S206* and *L3MBTL1*. Interestingly, the corresponding methylation associations were mostly reversed between the cell types for these three loci. Sites near *C15orf26* were all positively associated with telomere length in cell lines but negatively associated in whole blood cells. Sites near *L3MBTL1* and *RN5S206* tended to be negatively associated with telomere length in cell lines and positively associated in whole blood cells.

**Enrichment for positive associations near telomeres but not centromeres.** Of the 65 genes enriched for sites associated with telomere

**Table 1** | Identification of gene promoters enriched for sites at which methylation levels are associated with telomere length: summary data (SD = standard deviation, TSS = transcription start site)

|  | Blood (N = 24) | Cell-line (N = 36) |
|--|----------------|--------------------|
| Mean telomere length, T/S ratio (SD)           | 1.00 (0.29)    | 1.26 (0.45)        |
| Genes with TSS near associated CpG sites       | 65             | 36                 |
| Imprinted genes with associated CpG sites      | 12             | 3                  |
| Genes near telomeres with associated CpG sites | 12             | 8                  |
| Associated CpG sites (all)                     | 464            | 125                |
| Associated CpG sites (positive)                | 364 (78%)      | 94 (75%)           |
| Associated CpG sites (negative)                | 100 (22%)      | 31 (25%)           |



**Figure 2** | Heatmaps showing M-value variation for CpG sites (rows) across study participants (columns), in **A** whole blood and **B** cell-line DNA samples. Samples and CpG sites clustered using the Ward algorithm with Pearson's correlation as the distance metric. Relatively lower methylation levels are shown in green and higher methylation levels in red. The key for the telomere length of each sample in both heatmaps is given on the top right of the figure.

length in whole blood, 12 are located near telomeres (within 4 Mb of a chromosome end): *RP11-978I15.10*, *FAM50B*, *MAD1L1*, *CYP2E1*, *B4GALNT4*, *DRD4*, *CEND1*, *H19*, *KCNQ1OT1*, *CREBBP*, *RP11-1260E13.1* and *ADNP2*. Methylation levels in all but one of the CpG sites in these 12 loci (*RP11-1260E13.1*) are positively associated with telomere length, two examples of which are shown in Fig. 3. This represents a significant enrichment of associated genes located in subtelomeric regions ( $P < 0.01$ , Fisher's exact test (FET)). In contrast, there is no evidence for any enrichment of associated genes within 4 Mb of centromeres;  $P > 0.85$ , FET).

We observed nearly identical results in cell lines. Here, enrichment near telomeres was again significant for positively associated sites but not negatively associated sites ( $P = 0.012$  and  $P > 0.7$ , respectively, FET), and not for any associated sites near centromeres ( $P > 0.85$ , FET).

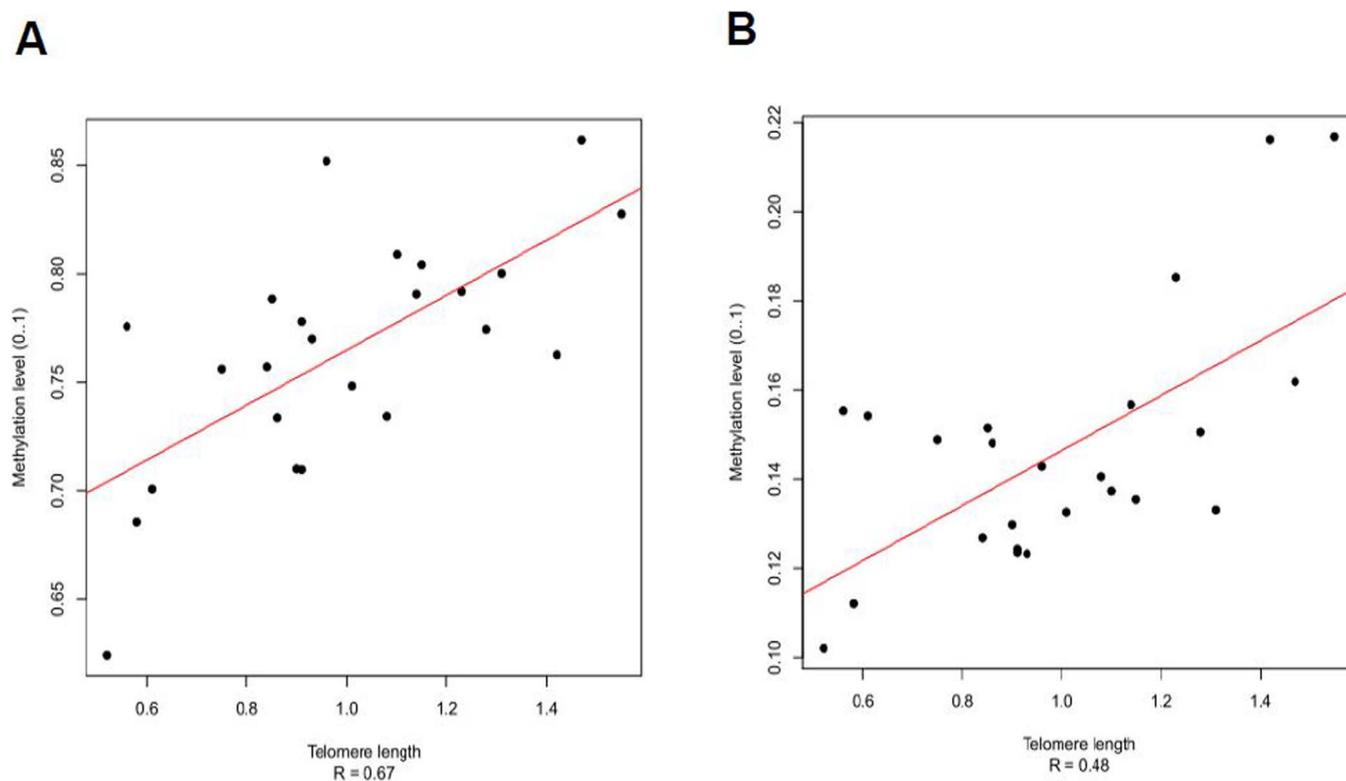
**Enrichment of imprinted genes.** Of the 65 genes with methylation levels associated with telomere length in whole blood DNA (Table 1), 12 are either confirmed (ten) or predicted (two) imprinted human genes, according to information in the genem印rint database (<http://www.genem印rint.com>, accessed January 2014). For seven of these loci (*MEST*, *KCNQ1OT1*, *DIRAS3*, *FAM50B*, *SGCE*, *L3MBTL1*, *GNAS-AS1*), only the paternal allele is expressed; for four (*H19*, *B4GALNT4*, *HOXA5*, *MEG3*) only the maternal allele is expressed; and for one (*BLCAP*) parental allele expression is isoform-dependent. Cross-checking of this information using a second database (<http://igc.otago.ac.nz/home.html>, accessed January 2014) also listed nine of the confirmed imprinted loci present in genem印rint, while the tenth (*DIRAS3*) was listed as provisional. Neither of the two loci listed as 'predicted' imprinted loci in genem印rint (*B4GALNT4* and *HOXA5*) were included in the second database. However, since there are only around 80 confirmed and 120 predicted imprinted loci in total in the human genome, the observation of even just ten of these genes in our study represents a highly significant enrichment ( $P < 1 \times 10^{-5}$ ; permutation test). Four of these imprinted loci (*FAM50B*, *H19*, *B4GALNT4* and *KCNQ1OT1*) are also located within 4 Mb of a telomere. Moreover, another gene for which methylation levels of nearby CpG sites was associated with telomere length is *ZFP57*, recently shown to be a global regulator of methylation in imprinting control regions (ICRs) in the genome<sup>35</sup>.

**Identification of potential ZFP57 binding sites in silico.** Following our discovery that telomere length is associated with methylation levels in both ICRs and the gene encoding a zinc finger protein responsible for their methylation status, *ZFP57*, we hypothesised that *Zfp57* may also play a role in the methylation status of subtelomeric regions. We therefore investigated whether subtelomeric regions are enriched for predicted *Zfp57* binding sites; a TGCCGC[A/G] sequence in which the cytosine of the CpG is methylated<sup>35</sup>. We identified 45,366 instances of this sequence (irrespective of methylation status) in the human genome (Build hg19/GRCh37) on both the forward and reverse strands. Of those, 8,547 are subtelomeric (within 4 Mb of the end of a chromosome) and 2,913 are centromeric (within 1.5 Mb from the centromere). If uniformly distributed, less than 6,000 motifs would be expected in subtelomeric regions, and more than 4,500 in centromeric regions. Hence, telomeres are enriched and centromeres are depleted of the predicted *Zfp57* binding motif ( $P < 2.2 \times 10^{-16}$ , hypergeometric). However, since the 450 K array only includes probes for 2,471 of the 45,366 predicted motifs, it was not possible to determine the methylation status of the majority of these target sequences, and thus their likely ability to bind *Zfp57* *in vivo*.

**Pathway analysis.** Table 2 shows the results of functional and network analyses carried out using the ingenuity pathway analysis system (IPA), for genes in which methylation levels of nearby CpG sites were associated with telomere length. These analyses were carried out separately for the results obtained in both blood and cell-line DNA. Of note, we observed significant enrichment of genes involved in endocrine disorders for the sites associated with telomere length in blood DNA, particularly diabetes mellitus (nine genes,  $P = 3.94e-4$ ) (Table S1). We also observed significant enrichment for imprinting disorders and various forms of cancers. The top canonical pathways enriched for associated loci were 'dopamine-DARPP32 feedback in cAMP signaling' in blood ( $P = 8.40e-03$ ) and 'B cell development' in cell-line DNA ( $P = 7.61e-04$ ).

## Discussion

Among 44–45 year-old men, we have identified multiple gene promoters enriched for sites at which methylation levels are associated with telomere length in human blood DNA. Furthermore, these associated loci are significantly overrepresented in subtelomeric



**Figure 3** | Scatterplots showing examples of positive associations between telomere length (T/S ratio, X-axis) in blood DNA and on the Y-axis, methylation levels at two loci located in subtelomeric regions: A) cg21024264, located 193 kb from chr10 qter within *CYP2E1* gene promoter region and B) cg25020933, located 369 kb from chr11 pter within *B4GALNT4* gene promoter region. Each point represents the mean T/S ratio of duplicate measurements for an individual DNA sample.

and imprinted genomic regions. We identified 65 gene promoters that contained sites associated with telomere length, which are largely distinct from the loci in which methylation levels have been associated with increasing age<sup>30–33</sup>. This lack of overlap is not unexpected, since our study examined telomere length variation at a single chronological age. However, four of the gene promoters that we found to be associated with telomere length also contained sites that belong to a set of 353 ‘age-predictor’ CpGs<sup>30</sup>: those at the *CYP2E1*, *DIRAS3*, *FAM50B* and *SGCE* loci. This may indicate that part of the epigenetic ‘signature’ of chronological age is related to telomere length shortening. It is well-known that there is wide inter-individual variation for the risk of age-related disease in people of the same chronological age. Loci at which methylation levels are associated with both chronological age *and* telomere length may thus be of particular relevance to the investigation of factors that influence successful ageing.

The associations identified in leukocyte DNA may represent either causal, consequential or coincidental relationships, that is, the promoters enriched for associated sites may be: i) genes that encode transcripts involved in regulating telomere length; ii) loci in which epigenetic changes are induced by changes in telomere length or iii) loci in which methylation levels are affected by cellular processes that also influence telomere length. We discuss our results in the context of each of these scenarios.

Several of the 65 loci enriched for CpG sites at which methylation levels are associated with telomere length have potential roles in human telomere biology, and are thus potential regulators of telomere length. Three of the positively-associated sites are within the *MAD1L1* gene, which has been shown in HeLa cells to act as negative regulator of *TERT*, the reverse transcriptase component of the telomerase enzyme<sup>36</sup>. Multiple positively-associated sites are located within the *POU5F1* locus, encoding the key pluripotency transcription factor

Oct-4, which regulates expression of the shelterin component TRF1 in mice<sup>37</sup>. We also identified both negative and positively associated sites within the *UI1* locus, which encodes part of the spliceosome, but is also essential for processing the RNA component of telomerase in fission yeast<sup>38</sup>. Furthermore, one of the negatively associated sites is located between the *MECOM* and *TERC* genes on chromosome 3; the latter encodes the RNA component of the human telomerase enzyme. Multiple common variants in this locus are associated with leukocyte telomere length in adults<sup>28</sup>. None of the other genes identified through GWAS of telomere length were amongst the loci for which we found methylation associations. However, we did find an association between methylation levels in the *MPL* locus and telomere length. A rare mutation in this gene causes a form of aplastic anaemia, a disorder also caused by mutations in the *TERC* and *TERT* genes<sup>39</sup>.

Associated sites may also be indicative of epigenetic changes induced by telomere length changes. We discovered that a significant proportion of the sites associated with telomere length are located in subtelomeric regions, and that for the majority of these, increased methylation levels are associated with longer telomeres. This finding supports previous evidence from animal and *in vitro* work that heterochromatin is lost in subtelomeric regions as telomeres shorten<sup>40</sup>. Age-related global hypomethylation of subtelomeric regions has been observed in both healthy Japanese individuals and those with Parkinson’s disease and sarcoidosis<sup>41–43</sup>, although the same group also report hypermethylation of subtelomeres in short telomeres in blood DNA from Alzheimer’s disease patients<sup>44</sup>. The authors postulate that this latter observation may result from the selective loss of cells with short, hypomethylated telomeres from the blood of Alzheimer’s patients. Further longitudinal studies of changes in telomere length and subtelomeric methylation levels in both healthy individuals and those affected by age-related disease are warranted to resolve this issue.



**Table 2 | Results of Ingenuity pathway and network analysis for genes with promoters enriched for CpG sites at which methylation levels are associated with telomere length in either blood DNA (65 genes) or cell-line DNA (36 genes). Details of the disorders and gene names for the analysis of blood DNA are given in Table S1**

| BLOOD DNA  |                     |           | CELL LINE DNA                                  |                     |           |
|--|---------------------|-----------|--|---------------------|-----------|
| <b>DISEASES AND DISORDERS</b>                        |                     |           |  |                     |           |
| Name   | P-value             | No. genes | Name   | P-value             | No. genes |
| Developmental Disorders                              | 9.59E-05 - 4.39E-02 | 13        | Developmental Disorders                        | 1.28E-03 - 4.40E-02 | 4         |
| Endocrine System Disorders                           | 9.59E-05 - 4.39E-02 | 18        | Endocrine System Disorders                     | 1.28E-03 - 3.91E-02 | 3         |
| Gastrointestinal Disease                             | 9.59E-05 - 4.77E-02 | 15        | Hereditary Disorders                           | 1.28E-03 - 4.77E-02 | 4         |
| Hereditary Disorders                                 | 9.59E-05 - 4.39E-02 | 15        | Infectious Disease                             | 1.28E-03 - 4.41E-02 | 6         |
| Reproductive System Disease                          | 9.59E-05 - 4.22E-02 | 14        | Metabolic Disease                              | 1.28E-03 - 3.91E-02 | 2         |
| <b>MOLECULAR AND CELLULAR FUNCTION</b>               |                     |           |  |                     |           |
| Name   | P-value             | No. genes | Name   | P-value             | No. genes |
| Gene Expression                                      | 1.86E-06 - 4.77E-02 | 19        | Cell Cycle                                     | 1.28E-03 - 3.66E-02 | 2         |
| Cellular Development                                 | 1.39E-04 - 4.77E-02 | 16        | Cellular Compromise                            | 1.28E-03 - 1.28E-03 | 1         |
| Cell Morphology                                      | 6.62E-04 - 3.68E-02 | 12        | Cellular Development                           | 1.28E-03 - 4.52E-02 | 8         |
| Cell-To-Cell Signaling and Interaction               | 1.07E-03 - 3.78E-02 | 10        | Cellular Movement                              | 1.28E-03 - 3.04E-02 | 4         |
| Drug Metabolism                                      | 1.07E-03 - 3.78E-02 | 3         | Energy Production                              | 1.28E-03 - 1.02E-02 | 2         |
| <b>PHYSIOLOGICAL SYSTEM DEVELOPMENT AND FUNCTION</b> |                     |           |  |                     |           |
| Name   | P-value             | No. genes | Name   | P-value             | No. genes |
| Connective Tissue Dev't and Function                 | 2.14E-05 - 4.03E-02 | 7         | Cardiovascular System Development and Function | 1.28E-03 - 4.15E-02 | 2         |
| Embryonic Development                                | 2.14E-05 - 4.81E-02 | 15        | Embryonic Development                          | 1.28E-03 - 4.52E-02 | 3         |
| Organ Development                                    | 2.14E-05 - 4.52E-02 | 14        | Nervous System Development and Function        | 1.28E-03 - 4.28E-02 | 3         |
| Organ Morphology                                     | 2.14E-05 - 4.52E-02 | 12        | Organ Development                              | 1.28E-03 - 4.52E-02 | 3         |
| Organismal Development                               | 2.14E-05 - 4.81E-02 | 13        | Organ Morphology                               | 1.28E-03 - 4.89E-02 | 6         |
| <b>TOP CANONICAL PATHWAYS</b>                        |                     |           |  |                     |           |
| Name   | P-value             | Ratio     | Name   | P-value             | Ratio     |
| Dopamine-DARPP32 Feedback in cAMP Signaling          | 8.40E-03            | 3 of 186  | B Cell Development                             | 7.61E-04            | 2 of 33   |
| Lymphotoxin b Receptor Signaling                     | 8.46E-03            | 2 of 61   | Nur77 Signaling in T Lymphocytes               | 2.24E-03            | 2 of 63   |
| Prod'n of NO and ROS in Macrophages                  | 1.08E-02            | 3 of 211  | IL-4 Signaling                                 | 4.01E-03            | 2 of 79   |
| Pregnenolone Biosynthesis                            | 1.53E-02            | 1 of 13   | SAPK/JNK Signaling                             | 6.53E-03            | 2 of 103  |
| Dopamine Receptor Signaling                          | 1.63E-02            | 2 of 95   | iCOS-iCOSL Signaling in T Helper Cells         | 8.21E-03            | 2 of 123  |

If confirmed, our finding that shorter telomeres are associated with decreased methylation levels of multiple cytosine sites located within 4 Mb of telomeres suggests a possible causal explanation for the relationship between shorter LTL and age-related diseases: as telomeres shorten, the resulting epigenetic changes in subtelomeric regions may alter the expression of disease-related genes. Such a mechanism was first postulated after the discovery of the 'telomere position effect' (TPE) in eukaryotic cells, in which the expression of transgenes located close to telomeres is repressed in a telomere length-dependent manner<sup>45</sup>. More recently, investigation of molecular pathology of facioscapulohumeral muscular dystrophy provides a precedent for the involvement of this mechanism in monogenic disease. Symptoms of this condition only appear once telomeres reach a critically short length. Stadler *et al* postulated that the effects of the mutant allele of the candidate gene concerned, *DUX4* (located in 4q35.2), are only apparent once it is "unmasked" by altered expression of this and other genes in subtelomeric regions<sup>46</sup>.

Several of the associated sites we identified in subtelomeric regions are within loci that have a known or potential role in age-related diseases. In particular, two positively associated sites are located within the *CREBBP* locus, which encodes a protein believed to play a central role in the pathogenesis of T2D<sup>47</sup>. Additionally, 15 positively associated sites are located within the *KCNQ1OT1* locus, also involved in T2D susceptibility<sup>48</sup>. Of potential relevance to cardiovascular disease are ten positively associated sites in *ADNP2*, which is involved in the cellular response to oxidative stress<sup>49</sup>.

Finally, we discovered a highly significant enrichment of telomere length-associated methylation sites near imprinted genes. We

speculate that this intriguing finding may indicate shared regulation of methylation status of both subtelomeric regions and imprinted genes – potentially via *ZFP57*, a locus in which we identified 21 sites positively associated with telomere length. This gene encodes a global regulator of imprinted genes, part of a specialised version of the KRAB-ZFP/KAP1 system recently been shown to be a general mechanism for establishing methylation patterns in the early mouse embryo. The *Zfp57* protein specifically targets only the methylated version of a TGCCGC[A/G] motif in ICRs, ensuring that methylation of the imprinted allele is maintained after fertilisation<sup>35</sup>. We determined that in addition to their expected presence within ICRs, there is enrichment of predicted *Zfp57* binding sites in subtelomeric regions. However, only 5.4% of the CpG sites within these predicted target sequences are detected by probes on the 450 K array used in our study. Thus, further investigations are required to determine the methylation status and actual binding of *Zfp57* within subtelomeric regions, in DNA isolated from different tissues and developmental stages.

We also investigated telomere length in DNA from EBV-transformed cell lines, which were created using the same venepuncture samples used to prepare the blood DNA samples. We identified multiple sites associated with cell-line telomere length, but there was minimal correlation between these and the sites associated with telomere length in blood DNA. This finding may reflect differences in the regulatory processes involved in controlling telomere length in somatic cells versus those involved in maintaining telomere length in immortalized cells. Telomere length in germ cells, haematopoietic and other stem cells is maintained by the action of telomerase.



Table 3 | Characteristics of the 38 male study participants

|  | Age (y) |              | n  |
|--|---------|--------------|----|
| Childhood manual social class n (%)            | 0       | 19 (50.0)    | 38 |
| Adulthood manual social class, n (%)           | 42      | 20 (52.6)    | 38 |
| Education level – below O level, n (%)         | 42      | 9 (23.7)     | 38 |
| Smoker, n (%)                                  | 42      | 10 (27.0)    | 37 |
| Alcohol drinks daily, n (%)                    | 42      | 9 (24.3)     | 37 |
| Birth weight, g, mean ± SD                     | birth   | 3522 (590)   | 37 |
| Height, m, mean ± SD                           | 7       | 1.23 (0.07)  | 35 |
| Height, m, mean ± SD                           | 42      | 1.77 (0.08)  | 37 |
| Body mass index, kg/m <sup>2</sup> , mean ± SD | 45      | 27.5 (4.07)  | 38 |
| Waist circumference, cm, mean ± SD             | 45      | 99.6 (10.5)  | 38 |
| Systolic blood pressure, mmHg, mean ± SD       | 45      | 134.4 (18.3) | 38 |
| Diastolic blood pressure, mmHg, mean ± SD      | 45      | 84.5 (11.6)  | 38 |
| FEV1*, mean ± SD                               | 45      | 3.82 (0.64)  | 34 |

\*FEV1 = one-second forced expiratory volume; best test of three spirometry readings.

However, a telomerase-independent pathway, known as “alternative lengthening of telomeres” (ALT), is activated in some cancer cells. There is also some evidence for activation of ALT in cells newly transformed with EBV<sup>50</sup>. This pathway synthesises telomeric sequences independently of telomerase, using homologous recombination following telomere sister chromatid exchange. Our results are consistent with distinct mechanisms being responsible for maintaining telomere length in blood and early passage EBV-transformed cells, with activation of the ALT pathway in transformed cells overriding DNA methylation states of genes that are associated with telomere length and telomerase activity in blood cells. Some of the CpG sites identified in our study as being associated with telomere length in cell-line DNA may provide further insights into the immortalisation process and the consequences of EBV infection.

Our study of methylation and telomere length was based on a relatively small number of samples, taken at one time point. Additionally, we only studied telomere length in blood DNA, so the relevance of our findings to other tissues remains to be determined. However, the LTL measurement method gives a mean value for the telomeres of all chromosomes in all the different types of leukocyte present in blood, which in turn is presumed to reflect telomere length of the haematopoietic stem cells. It is likely that mechanisms that regulate telomere length are common to all healthy somatic tissues and the stem cells from which they are derived, since there is high correlation between telomere length in multiple tissues from the same individuals<sup>51</sup>.

One strength of our study is the homogenous nature of the population with respect to age and gender (i.e. all men aged 44–45 at the time blood samples were taken). Accordingly, the potential confounding effects of these variables - both known to be strongly associated with telomere length and DNA methylation - were absent from our analyses. A further consequence of the age group examined is that our findings represent methylation associations with telomere length, rather than chronological age, and are thus of relevance to the study of healthy ageing.

The results of our pilot study suggest that as telomeres shorten, the methylation levels of many gene promoters in subtelomeric regions may change, which in turn could cause changes in gene expression that increase the risk of age-related disease. Since LTL is highly heritable, this supports the notion that ‘telotype’ may contribute towards explaining some of the missing heritability observed for many common conditions with a genetic component, as proposed by Armanios and Blackburn<sup>52</sup>. Further functional and epidemiological investigations of these observations are required to confirm the role of telomere length-dependent gene expression in health and disease.

## Methods

**Samples.** Genomic DNA samples from both whole blood and EBV cell-lines from adult males aged 44–45, participants in the 1958 British Birth Cohort Study, were available for this study<sup>53</sup>. Details of these samples have been reported previously<sup>54</sup>, and the characteristics of the 38 participants included in this study are summarised in Table 3. In brief, 40 male participants were originally selected to represent extremes of socioeconomic position in child and adulthood, and other characteristics namely, *in utero* exposure to tobacco and childhood abuse. Telomere length measurements and DNA methylation data were obtained using the same DNA sample for 38 of the original 40 males (those for whom enough blood or cell-line DNA was available for both these analyses). DNA from whole blood was available for 24 males and from EBV cell lines N = 36, while for 22 participants both whole blood and EBV cell line DNA was available.

**Ethics statement.** All participants provided written consent and a blood sample for DNA analysis and EBV-transformation into a lymphoblastoid cell-line for future studies. Ethical approval was given by the South-East Multi-Centre Research Ethics Committee. All methods were carried out in accordance with approved guidelines.

**Telomere measurements.** Mean relative telomere length was measured in genomic DNA samples prepared from either whole blood or EBV cell-lines, using a multiplex quantitative real-time PCR method<sup>34</sup> with minor modifications as described previously<sup>22</sup>. To minimise intra-assay variability, all PCRs were carried out on a single 384-well plate on a CFX384 Real-time PCR detection system (Bio-Rad). Five serial dilutions of a reference sample (leukocyte DNA from a 42-year-old non-cohort female) spanning 5–50 ng were run in triplicate, in addition to a no-template control (NTC). Human beta-globin (Hgb) was used as the single copy reference gene. Following amplification and data collection, the CFX manager software (Bio-rad) was used to generate standard curves for the reference DNA dilutions, one for the telomere signal (T) and one for the single copy gene signal (S). Telomere measurements for each sample were calculated as T/S ratios, a relative measure of the amplification of the telomeric DNA sequence compared to that of the single copy gene. The T/S values reported are the mean of duplicate measurements for each sample, and the overall mean coefficient of variation (CV) between duplicates was 4%.

**Infinium HumanMethylation450 microarray analysis.** DNA methylation profiles were generated from the DNA of whole blood samples and EBV cell lines using the Infinium HumanMethylation450 microarray<sup>55</sup>. In all, there were 22 study participants with both a whole blood DNA methylation profile as well as an EBV cell line methylation profile.

Microarray quality was assessed by generating quality control plots using the minfi Bioconductor package and found to be satisfactory<sup>56</sup>. Microarrays were background corrected and normalised to control probes using the minfi package. The probe intensities were then transformed into beta-values ( $M/(M + U)$ ) where M and U are the corresponding methylated, and unmethylated, probe intensities, respectively. Although beta-values can be conveniently interpreted as methylation levels, association analyses were carried out using so-called M-values, logit-transformed beta values as previously recommended<sup>57</sup>.

Statistically significant associations between DNA methylation and telomere length were identified by fitting linear models using the limma Bioconductor package<sup>58</sup> with microarray M-values as the dependent variables and sample telomere lengths as the independent. In addition, models included adjustments for potential confounding due to technical artifacts such as plate effects, as well as individual characteristics including childhood and adulthood SEP, childhood abuse and maternal smoking during pregnancy. Adjustments were made by including independent surrogate variables in the limma design matrix as identified by the DoISVA function of the isva R package<sup>59</sup>. The DoISVA function was applied to the M-values twice, once with no potential confounders included in the input and a second time with the technical and individual variables noted above along with any statistically significant variables identified in the first call to DoISVA (False discovery rate (FDR) < 0.05). The resulting two most statistically significant independent surrogate variables were then retained for inclusion in the limma design matrix.

In order to identify genes likely to be affected by telomere length-associated methylation, each CpG represented on the microarray was assigned the Ensembl gene identifier with the nearest TSS not more than 10 Kb away (retrieved January 2013). The Wilcoxon rank-sum test was then applied to compare the limma t-statistics of the CpG sites associated with each gene to the limma t-statistics of the CpG sites not associated with the gene. Hence, a significant P-value indicates that the CpG sites associated with the gene are enriched for methylation levels either positively or negatively correlated with telomere length. FDRs were calculated from these P-values in order to control for multiple testing using the Benjamini-Hochberg algorithm. The methylation levels of an individual CpG site were considered significantly associated with telomere length if the FDR of its associated gene was < 0.05 (“Q value”) and if the unadjusted P value for the site calculated by limma was < 0.05. All such sites are listed in Data S1 and S2.

**Enrichment of sites near imprinted genes.** Given how methylation associations are identified, it is not possible to simply apply Fisher’s exact test to determine whether there is a surprisingly large number of imprinted genes near significantly associated CpG sites. This is because genes near a large number of CpG sites represented on the microarray are more likely to be near a significantly associated CpG site than another gene near a small number of CpG sites. We confirmed that this was the case for



imprinted genes, which are on average near twice as many microarray CpG sites as expected. Consequently, we designed a permutation test that adjusts for this bias. We first identified the nearest gene TSS for each CpG site and then calculated for each gene the number of such CpG sites, called the *nearest count* of the gene. For our permutation test, we repeatedly selected (100,000 times) gene sets with the same size and “nearest count” distribution as the set of genes found significantly associated with telomere length. Whereas our associated gene set contained 12 imprinted genes, the average random gene set contained only 1 imprinted gene, and only one random gene set contained 10 imprinted genes.

**Pathway analysis.** Functional and network relationships between the genes associated with CpG sites in which methylation levels are correlated with telomere length were investigated using the Ingenuity Pathway Analysis (IPA) software tool (Ingenuity Systems, www.ingenuity.com).

- de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
- von Zglinicki, T. Oxidative stress shortens telomeres. *TIBS* **27**, 339–344 (2002).
- Kurz, D. J. *et al.* Chronic oxidative stress compromises telomere integrity and accelerates the onset of senescence in human endothelial cells. *J. Cell Sci.* **117**, 2417–2426 (2004).
- Aviv, A. *et al.* Menopause modifies the association of leukocyte telomere length with insulin resistance and inflammation. *J. Clin. Endocrinol. Metab.* **91**, 635–640 (2006).
- von Zglinicki, T., Saretzki, G., Docke, W. & Lotze, C. Mild hyperoxia shortens telomeres and inhibits proliferation of fibroblasts: a model for senescence? *Experimental Cell Research* **220**, 186–193 (1995).
- Hemann, M. T., Strong, M. A., Hao, L. Y. & Greider, C. W. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* **107**, 67–77 (2001).
- Fitzpatrick, A. L. *et al.* Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am. J. Epidemiol.* **165**, 14–21 (2007).
- Kimura, M. *et al.* Telomere length and mortality: a study of leukocytes in elderly Danish twins. *Am. J. Epidemiol.* **167**, 799–806 (2008).
- Oeseburg, H., de Boer, R. A., van Gilst, W. H. & van der Harst, P. Telomere biology in healthy aging and disease. *Pflug. Arch. Eur. J. Phy.* **459**, 259–268 (2010).
- Mather, K. A., Jorm, A. F., Parslow, R. A. & Christensen, H. Is telomere length a biomarker of aging? A review. *J. Gerontol. Ser. A-Biol. Sci. Med. Sci.* **66**, 202–213 (2011).
- Bakaysa, S. L. *et al.* Telomere length predicts survival independent of genetic influences. *Aging Cell* **6**, 769–774 (2007).
- Austad, S. N. Why women live longer than men: sex differences in longevity. *Gender Med.* **3**, 79–92 (2006).
- Brouillette, 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. *Lancet* **369**, 107–114 (2007).
- Adaikalakoteswari, A., Balasubramanyam, M. & Mohan, V. Telomere shortening occurs in Asian Indian Type 2 diabetic patients. *Diabet. Med.* **22**, 1151–1156 (2005).
- Salpea, K. *et al.* Association of telomere length with type 2 diabetes, oxidative stress and UCP2 gene variation. *Atherosclerosis* **209**, 42–50 (2009).
- Prescott, J., Wentzensen, I. M., Savage, S. A. & De Vivo, I. Epidemiologic evidence for a role of telomere dysfunction in cancer etiology. *Mutat. Res.* **730**, 75–84 (2012).
- Okuda, K. *et al.* Telomere Length in the Newborn. *Pediatr. Res.* **52**, 377–381 (2002).
- 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).
- Valdes, A. M. *et al.* Obesity, cigarette smoking, and telomere length in women. *Lancet* **366**, 662–664 (2005).
- Strandberg, T. E. *et al.* Association between alcohol consumption in healthy midlife and telomere length in older men. The Helsinki Businessmen Study. *Eur. J. Epidemiol.* **27**, 815–22 (2012).
- Epel, E. S. *et al.* Accelerated telomere shortening in response to life stress. *Proc. Natl Acad. Sci. USA* **101**, 17312–17315 (2004).
- Buxton, J. L. *et al.* Childhood Obesity Is Associated with Shorter Leukocyte Telomere Length. *J. Clin. Endocrinol. Metab.* **96**, 1500–1505 (2011).
- Ala-Mursula, L. *et al.* Long-term unemployment is associated with short telomeres in 31-year-old men: an observational study in the northern Finland birth cohort 1966. *PLoS ONE* **8**, e80094 (2013).
- Tyrka, A. R. *et al.* Childhood Maltreatment and Telomere Shortening: Preliminary Support for an Effect of Early Stress on Cellular Aging. *Biol. Psychiatry* **67**, 531–534 (2010).
- Shalev, I. *et al.* Exposure to violence during childhood is associated with telomere erosion from 5 to 10 years of age: a longitudinal study. *Mol. Psychiatry*, **18**, 576–81 (2012).
- 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).
- Mangino, M. *et al.* Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Hum. Mol. Genet.* **21**, 5385–5394 (2012).
- Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.* **45**, 422–427 (2013).
- Chen, L. Y., Redon, S. & Lingner, J. The human CST complex is a terminator of telomerase activity. *Nature* **488**, 540–4 (2012).
- Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 [Epub ahead of print] (2013).
- McClay, J. L. *et al.* A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Hum. Mol. Genet.* [Epub ahead of print] (2013).
- Heyn, H. *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc. Natl Acad. Sci. USA* **109**, 10522–10527 (2012).
- Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
- Cawthon, R. M. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res.* **37**, e21 (2009).
- Quenneville, S. *et al.* The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell Rep.* **2**, 766–773 (2012).
- Lin, S. Y. & Elledge, S. J. Multiple tumor suppressor pathways negatively regulate telomerase. *Cell* **113**, 881–889 (2003).
- Schneider, R. P. *et al.* TRF1 is a stem cell marker and is essential for the generation of induced pluripotent stem cells. *Nat. Commun.* **4**, 1946 (2013).
- Box, J. A., Bunch, J. T., Tang, W. & Baumann, P. Spliceosomal cleavage generates the 3' end of telomerase RNA. *Nature* **456**, 910–914 (2008).
- Walne, A. J. *et al.* Exome sequencing identifies MPL as a causative gene in familial aplastic anemia. *Haematologica* **97**, 524–528 (2012).
- Benetti, R., Garcia-Cao, M. & Blasco, M. A. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat. Genet.* **39**, 243–250 (2007).
- Maeda, T., Guan, J. Z., Higuchi, Y., Oyama, J. & Makino, N. Aging-related alterations of subtelomeric methylation in sarcoidosis patients. *J. Gerontol. Ser. A-Biol. Sci. Med. Sci.* **64**, 752–760 (2009).
- Maeda, T., Guan, J. Z., Oyama, J., Higuchi, Y. & Makino, N. Aging-associated alteration of subtelomeric methylation in Parkinson's disease. *J. Gerontol. Ser. A-Biol. Sci. Med. Sci.* **64**, 949–955 (2009).
- Maeda, T., Guan, J. Z., Oyama, J., Higuchi, Y. & Makino, N. Age-related changes in subtelomeric methylation in the normal Japanese population. *J. Gerontol. Ser. A-Biol. Sci. Med. Sci.* **64**, 426–434 (2009).
- Guan, J. Z., Guan, W. P., Maeda, T. & Makino, N. The subtelomere of short telomeres is hypermethylated in Alzheimer's Disease. *Aging Dis.* **3**, 164–70 (2012).
- Baur, J. A., Zou, Y., Shay, J. W. & Wright, W. E. Telomere position effect in human cells. *Science*. **292**, 2075–7 (2001).
- Stadler, G. *et al.* Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat. Struct. Mol. Biol.* **20**, 671–678 (2013).
- Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
- Travers, M. E. *et al.* Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets. *Diabetes* **62**, 987–992 (2013).
- Kushnir, M., Dresner, E., Mandel, S. & Gozes, I. Silencing of the ADNP-family member, ADNP2, results in changes in cellular viability under oxidative stress. *J. Neurochem.* **105**, 537–545 (2008).
- Kamranvar, S. A., Chen, X. & Masucci, M. G. Telomere dysfunction and activation of alternative lengthening of telomeres in B-lymphocytes infected by Epstein-Barr virus. *Oncogene* **32**, 5522–5530 (2013).
- Daniali, L. *et al.* Telomeres shorten at equivalent rates in somatic tissues of adults. *Nat. Commun.* **4**, 1597 (2013).
- Armanios, M. & Blackburn, E. H. The telomere syndromes. *Nat. Rev. Genet.* **13**, 693–704 (2012).
- Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int. J. Epidemiol.* **35**, 34–41 (2006).
- Borghol, N. *et al.* Associations with early-life socio-economic position in adult DNA methylation. *Int. J. Epidemiol.* **41**, 62–74 (2012).
- Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–295 (2011).
- Minfi: Analyze Illumina's 450 k Methylation Arrays (Vienna: R Foundation for Statistical Computing, 2011).
- Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
- Smyth, G. K. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds Gentleman, R. *et al.*) 397–420 (Springer, 2005).
- Teschendorff, A. E., Zhuang, J. & Widschwendter, M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics* **27**, 1496–1505 (2011).

## Acknowledgments

The Section of Investigative Medicine at Imperial College London is funded by grants from the MRC, BBSRC, NIHR, an Integrative Mammalian Biology (IMB) Capacity Building Award, and FP7- HEALTH- 2009- 241592 EuroCHIP grant and is supported by the NIHR



Imperial Biomedical Research Centre Funding Scheme. For the 1958 Birth Cohort samples, venous blood collection was funded by the UK Medical Research Council (G0000934) and cell-line production, DNA extraction and processing by the Wellcome Trust (068545/Z/02). MS was supported by a grant from the Canadian Institute for Health Research MOP-42411. This work was also funded by a Wellcome Trust programme grant 072937/Z/03/Z awarded to MP, and a Wellcome Trust fellowship grant, WT088431MA, held by JLB.

### Author contributions

J.L.B. performed telomere measurements, contributed to data analysis, and co-wrote the manuscript, M.S. carried out data analysis and co-wrote the manuscript, J.P. and N.B. generated DNA methylation data and contributed to data analysis, WM contributed DNA samples, A.I.F.B., C.H. (Deceased February 8, 2013), C.P., M.S. and M.P. contributed to writing of the manuscript and study design, and M.P. additionally supervised the study.

### Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Buxton, J.L. *et al.* Human leukocyte telomere length is associated with DNA methylation levels in multiple subtelomeric and imprinted loci. *Sci. Rep.* **4**, 4954; DOI:10.1038/srep04954 (2014).



This work is licensed under a Creative Commons Attribution 3.0 Unported License. The images in this article are included in the article's Creative Commons license, unless indicated otherwise in the image credit; if the image is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the image. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>