

OPEN

WGS-based telomere length analysis in Dutch family trios implicates stronger maternal inheritance and a role for *RRM1* gene

Lilit Nersisyan^{1*}, Maria Nikoghosyan^{1,2}, The Genome of the Netherlands consortium[†] & Arsen Arakelyan^{1,2}

Telomere length (TL) regulation is an important factor in ageing, reproduction and cancer development. Genetic, hereditary and environmental factors regulating TL are currently widely investigated, however, their relative contribution to TL variability is still understudied. We have used whole genome sequencing data of 250 family trios from the Genome of the Netherlands project to perform computational measurement of TL and a series of regression and genome-wide association analyses to reveal TL inheritance patterns and associated genetic factors. Our results confirm that TL is a largely heritable trait, primarily with mother's, and, to a lesser extent, with father's TL having the strongest influence on the offspring. In this cohort, mother's, but not father's age at conception was positively linked to offspring TL. Age-related TL attrition of 40 bp/year had relatively small influence on TL variability. Finally, we have identified TL-associated variations in ribonuclease reductase catalytic subunit M1 (*RRM1* gene), which is known to regulate telomere maintenance in yeast. We also highlight the importance of multivariate approach and the limitations of existing tools for the analysis of TL as a polygenic heritable quantitative trait.

The terminal regions of linear human chromosomes are composed of telomeres - sequences of tandem TTAGGG repeats normally stretching around 10–15 kb in length. They protect the chromosomes from degradation and end-to-end fusions¹ and perform a number of regulatory functions, including regulation of gene expression², DNA damage response³, modulation of cellular senescence⁴, proper chromosome anchoring to the nuclear membrane and segregation during meiosis⁵. Accordingly, changes in telomere length (TL) have been linked to development of age-related diseases and cancers^{6–9}.

TL regulation is a complex phenomenon that depends on a number of variables, such as age¹⁰, genetic^{11,12} and environmental^{13,14} factors, and the tissue under study^{15,16}. Initially, the idea of telomere length dependence on chronological age was based on the fact that in most somatic cells, telomeres get shorter over an individual's lifespan, mostly due to telomere end-resection that occurs after each round of DNA replication^{17,18}. However, while telomere attrition is associated with cellular senescence, it's not completely clear how TL is linked to organismal ageing¹⁹. Particularly, inverse correlation of TL with age has been shown to be true for some^{20,21}, but not for all populations^{22,23}. Our previous study on South Asian genomes has identified no relationship with age in this population²³, confirming previous observations suggesting that age-dependency of telomere length might be population-specific²².

Genetic factors influencing TL have been identified by multiple genome-wide association studies, both in healthy and in diseased populations^{11,12}. Single nucleotide variations (SNVs) in several loci, such as telomerase components *TERT* and *TERC*, and others (*OBFC1*, *RTEL1*, etc.) have been linked to TL^{11,24,25}. For many loci, however, the identified associations with telomere length are either irreproducible in other cohorts or show

¹Bioinformatics Group, Institute of Molecular Biology NAS RA, 7 Hasratyan str., 0014, Yerevan, Armenia. ²Institute of Biomedicine and Pharmacy, Russian-Armenian University, 123 Hovsep Emin St, 0051, Yerevan, Armenia. [†]A comprehensive list of consortium members appears at the end of the paper. *email: l_nersisyan@mb.sci.am

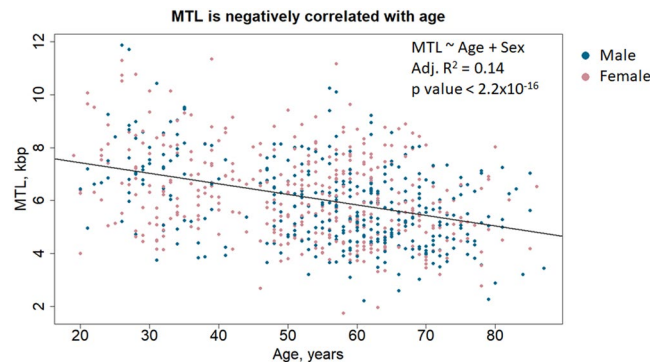


Figure 1. Negative association of telomere length with age. Combined analysis of children and parents shows ~40 bp reduction in MTL per year. Sex was not significantly associated with MTL after adjustment for age. Overall, age explains only 14% of MTL variability among individuals.

population-specificity²⁶. Additionally, a number of environmental factors, such as stress, exposure to oxidative agents, smoking, alcohol consumption and lifestyle, all affect telomere attrition rate^{13,14}.

Finally, while telomere length regulation and environmental factors affect TL variability in a lifetime of a single individual, TL variability between individuals is also largely explained by inheritance of parental telomeres and variable TLs at birth, with a stronger link between offspring and either father's²⁷, or mother's²⁸ TL. Additionally, paternal (PAC) and maternal (MAC) ages at conception also affect TL in the offspring²⁹. This has mostly been shown for PAC, and explained by relatively longer TL in the sperm of older males^{27,30,31}. A recent study has simulated this effect across many generations and has suggested this as a mechanism for adaptation to cancer³².

Overall, based on their importance in healthy ageing and in regulation of basic cellular functionality, telomeres are being extensively studied for their role in development of age-related diseases, cancers and reproduction^{6,10,24,25}. Involvement of multiple genetic and environmental factors and heritability of TL largely complicates investigation of TL regulation and necessitates multi-factor analysis. We have previously studied TL in the South Asian population, based on whole genome sequencing (WGS) data. In this population, we have identified TL-associated loci in the *ADARB2* gene, but did not reveal relation of TL with age²³. That dataset, however, did not allow for analysis of TL inheritance patterns. Here, we make use of whole-blood derived DNA sequencing data generated by the Genome of the Netherlands (GoNL) project that involves WGS and SNV datasets from 250 families, comprising trios of a mother, a father and a child. Importantly, this dataset was annotated not only for ages of individuals at the time of data collection, but also for parental ages at conception³³. We have measured mean telomere length (MTL) from WGS data using Computel³⁴, and have performed a number of regression analyses to identify age association and inheritance patterns of MTL in the Dutch population, and to reveal MTL-associated genomic loci.

Results

Population stratification. Principal component analysis on SNVs was previously performed by the GoNL consortium³³. They had identified a subtle substructure of the population according to geographical origin of the individuals (along the north-south axis of the Netherlands). However, there were no distinct clusters and the first and the second principal components explained very little variation in the genotype data (~0.25% and ~0.24%, respectively). Therefore, we have performed downstream association analyses with the assumption of a homogeneous population.

Regression analysis. In the studied population, MTL ranged from as low as 1.7 to 12 kbp in length. Multiple linear regression (MLR) analysis (with the model $MTL \sim Age + Sex$) on all study subjects, including the MTL of parents and children, has shown that age is negatively correlated with MTL (adjusted $R^2 = 0.14$, p value $< 2.2e-16$), with an estimated attrition rate of ~40 bp per year, in agreement with prior observations^{10,35} (Fig. 1). On the other hand, contrary to studies commonly reporting longer telomeres in adult women compared to men³⁶, in our cohort sex was not a significant predictor of MTL. Even though females on average had longer telomeres in our population (5.9 ± 1.6 kb in females versus 5.7 ± 1.6 kb in males, t -test p value = 0.03), this is likely due to age differences, as women were generally younger than men (by 4 years, t test p value < 0.001). Indeed, this difference was gone after adjustment for age by testing the mean difference in MTLs between males and females in each age group (females on average had 200 bp longer MTL, but with t -test p value 0.24).

Next, we have performed a series of MLRs to analyze the inheritance patterns of maternal and paternal telomeres, using MTL of children as an outcome variable and MTL of parents as predictors. For each child in our dataset, we had information regarding their age and MTL, maternal MTL (mMTL), paternal MTL (fMTL), as well as mother's and father's ages at conception (MAC and PAC). Pairwise correlation analysis for all these predictors (Supplementary Fig. S1) have shown that many variables were not independent, with mother's and father's, as well as with parents' and child's ages showing strong correlation, and MTL, mMTL and fMTL being correlated with each other. In order to avoid collinearity related biases, we have constructed and tested simple and complex models including one to many predictors, as well as interaction terms. Among the models with highest adjusted R^2 value, were the simple model $MTL \sim Age + mMTL + fMTL + MAC + PAC$ (adj. $R^2 = 0.442$, Supplementary

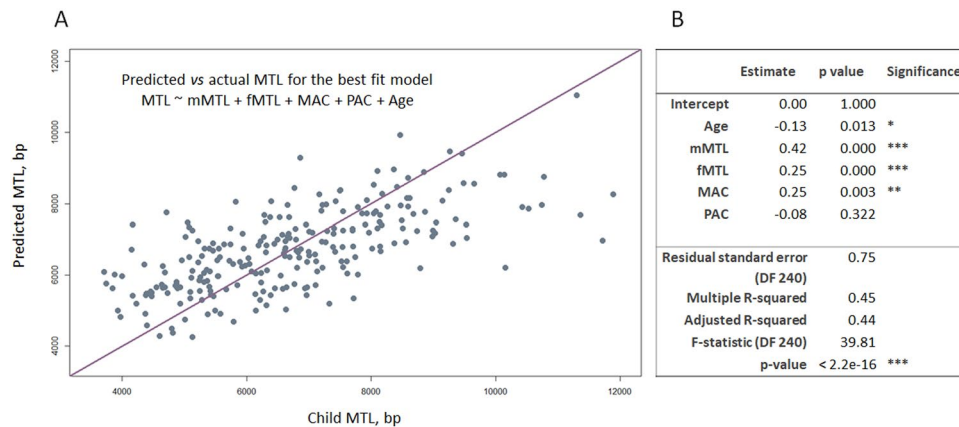


Figure 2. Multiple linear regression estimates for the best fit model of telomere length inheritance. **(A)** The actual MTL length of the offspring and the MTL predicted with the best fit model of telomere length inheritance: $MTL \sim Age + mMTL + fMTL + MAC + PAC + Age$. **(B)** The correlation estimates, and p values for MTL with offspring age, maternal and paternal MTL (mMTL, fMTL) and maternal and paternal ages at conception (MAC, PAC) are presented, along with general statistics for the test (bottom). Significant correlations are marked with one, two and three asterisks, for p values of < 0.05 , < 0.01 and < 0.001 , respectively. Note, that PAC was not a significant predictor in the model: its estimates are presented for the reader's information.

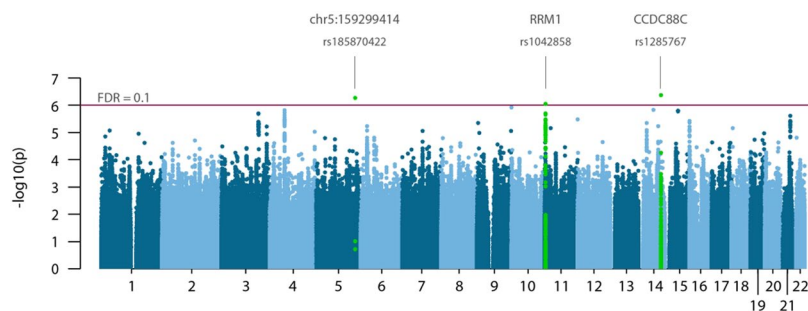


Figure 3. Manhattan plot for significance of SNV association with MTL. Analysis were performed with Plink family based association tests, with additive model for SNVs. Three SNVs (rs185870422 – chromosome 5q, rs1042858 – RRM1, rs1285767 – CCDC88C) are significantly associated with MTL at false discovery rate of 0.1.

Table S1), and two complex models including interaction terms. As the goodness of fit and predictors' correlation coefficients were quite close to each other among these models, we have chosen the simple one for interpretability (see Supplementary Table S1). Exploration of regression estimates for each predictor shows that age is correlated with a scaled estimate of -0.34 when it's the only predictor in the model. However, its influence gets three times weaker after inclusion of mother's and father's MTLs, which are stronger predictors of child's MTL (single-model estimates 0.58 and 0.51, respectively).

The final model explained only 44% of the variability in the offspring MTLs (Fig. 2A), with the strongest correlate being mother's MTL (correlation estimate = 0.42), and to a lesser extent – father's (correlation estimate = 0.25). In this model, age was still negatively correlated with MTL, however with smaller regression estimate of only -0.13 (Fig. 2B). Interestingly, MAC was significantly and positively correlated with child's MTL (0.25), while PAC was not a significant predictor. The observed association between MAC and child's MTL still preserved after accounting for the rest of the variables with partial correlation analysis (partial correlation estimates for the association with offspring MTL after adjustment for the rest of the factors were -0.16 (Age, $p < 0.05$), 0.42 (mMTL, $p < 0.001$), 0.26 (fMTL, $p < 0.001$), 0.19 (MAC, $p < 0.01$) and -0.06 (PAC, $p > 0.05$).

Telomere length qTL analysis. Family based quantitative trait association analyses with MTL were performed with Plink and Merlin. The lists of identified top SNVs are available in Supplementary Datasheet S1. Plink revealed three SNVs significantly associated with telomere length ($FDR < 0.1$): rs1285767 (CCDC88C), rs185870422 (non-coding region on chromosome 5) and rs1042858 (RRM1) (Fig. 3). The current version of Plink (version 1.9) did not allow for inclusion of additional covariates, namely age, in the family based MTL association analysis. Thus, we have also performed family association tests with Merlin to obtain independent assessments of qTL. Merlin did account for age as a covariate, however, did not account for mendelian heritability of MTL. Thus, no SNVs were identified as significantly associated with MTL with this test. This is in accordance with MTL inheritance patterns revealed by MLR analysis described above, as telomere inheritance factors, particularly

Gene (SNV)	CCDC88C rs1285767	chr 5q rs185870422	RRM1 rs1042858	DSCAM rs62237635 & six other linked SNVs
Association test				
Plink, family association	4.30E-07	5.50E-07	9.00E-07	0.066
Plink, regression parents	1.02E-05	0.002	4.68E-05	> 0.05
Plink, regression children	0.018	0.0007	> 0.05	9.29E-08
Merlin, family association	> 0.05	NA (not coding)	> 0.05	0.007
MLR inheritance + SNV	0.011	0.004	> 0.05	1.15E-07

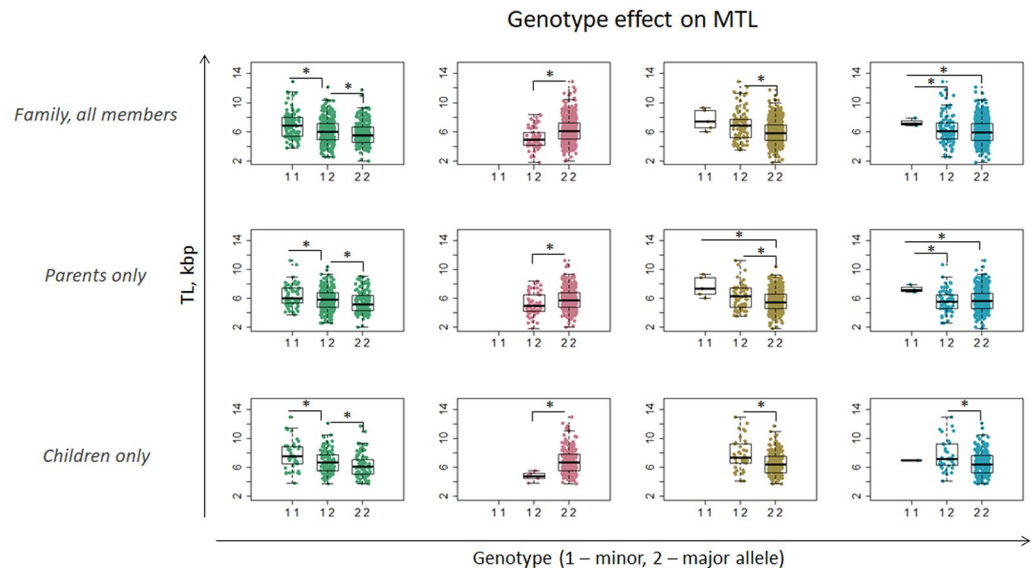


Figure 4. Significance and genotype effects of MTL-associated SNVs. Colors are used to delineate each SNV. The dark-to-light color gradient indicates low-to-high p values (raw p values from each test). The value of 1×10^{-7} (E-07) corresponds to FDR threshold of 0.1. Family tests with Merlin only accounted for coding regions, therefore no p-value for the SNV at chr5q was available for this test. The DSCAM gene was identified with six SNVs (rs11701674, rs28412810, rs11700591, rs62237637, rs11702319), which, however, were linked and showed similar patterns represented here with rs11701674 only. Boxplots show MTL values for each individual in the genotype. Significant MTL differences between genotypes are highlighted with asterisks. All the SNVs show consistent effect in parents and in children, with the minor allele of SNV at chr5q having negative, while that of the other three SNVs having positive association with MTL.

mother's and father's MTL and MAC (correlation coefficients 0.42, 0.25 and 0.25), are stronger predictors of MTL than age (correlation coefficient -0.13).

Taking into consideration the limitations of these two methods, we have also performed multiple linear regression tests for each SNV, by adding child's genotype for each SNV to the MLR model obtained above: $MTL \sim Age + mMTL + fMTL + MAC + PAC + SNV$. Here each SNV takes on values from 0 to 2, depending on the number of dominant alleles in a child's genotype. These tests revealed only six SNVs residing within the DSCAM gene, associated with offspring MTL after adjustment for all the variables included in the MLR model. With this model, only two top SNVs from Plink association tests had significant association p values (unadjusted) (rs185870422 (chr5q) and rs1285767 (CCDC88C)). The results for these two SNVs and DSCAM were further confirmed by Plink linear regression test performed on the population of children only. Isolated Plink linear regression tests on the parents' population did not reveal any significant quantitative loci after multiple comparison test adjustment (Fig. 4, Supplementary Datasheet S1).

One of the three loci identified by Plink, rs1285767, resides within the CCDC88C gene and leads to 514 bp gain of MTL per additional minor allele. CCDC88C encodes the protein Daple that is known to regulate the non-canonical Wnt signaling pathway, which plays an important role in embryonic development, cell motility and tumor progression^{37,38}. Wnt signaling pathway is a known regulator of telomerase activity through its central player β -catenin, which binds transcription factors and regulates expression of TERT, the catalytic subunit of telomerase³⁹. Nonetheless, the SNV is positioned intronically and its functional effects are unknown. This locus came out as a significant MTL-associated hit not only in Plink family association analysis, but was also border-line significant in separate Plink regression analyses on the populations of parents and children. The presence of minor alleles shows consistent additive effects on MTL in family and parent/children groups (Fig. 4).

The next locus identified by Plink-based family tests, as well as by Plink linear regression analysis on parent/children groups was rs185870422, residing in a non-coding region on the q arm of chromosome 5. Interestingly, no individual had a homozygous genotype for the minor allele of this SNV. However, the heterozygotes had consistently lower MTL (by 1000 bp) than homozygotes for the major allele (Fig. 4). This locus distantly interacts with *UBLCP1* and *RNF145* genes within the same chromatin loop (source: database 3DSNP version 1.0). Those, however, do not have known roles in telomere length regulation.

The *RRM1* locus was the most consistent hit in Plink family tests, with 75 SNVs having border-line significance (raw p value $< 10^{-6}$ – 10^{-5}), in contrast to *CCDC88C*, which had only two SNVs with raw p values $< 10^{-5}$ (Fig. 4, Supplementary Datasheet S1). The rs1042858 variant of *RRM1* had positive effect on MTL (892 bp per locus), but, it was only identified during Plink family association tests and separate regression tests on parents (Fig. 4). Homozygotes for the minor allele were missing among children, but the presence of the minor allele in heterozygotes had the same positive effect on MTL. According to Ensembl Variant Effect Predictor results, this variant is synonymous to the standard transcript of *RRM1*, but can, however, introduce a modification at the 3' UTR of its alternative transcript (ENST00000532170.5) that may affect nonsense-mediated decay.

The six SNVs revealed by the MLR model tests (rs11701674, rs28412810, rs11700591, rs62237637, rs11702319) were all linked and resided within intronic regions of the Down Syndrome cell adhesion molecule (*DSCAM*) gene, which encodes an immunoglobulin superfamily member, is involved in axon guidance and is a candidate for Down syndrome and congenital heart disease⁴⁰. The amount of homozygotes for the minor alleles of these SNVs was low, and their effect on TL was only seen in children. In the MLR inheritance model, the regression coefficient for these SNVs was higher (0.67) compared to all other variables (mMTL had a correlation coefficient of 0.42 after addition of these SNVs). However, addition of these SNVs to the MLR model increased the percentage of explained variation in offspring MTL by only 6% (leading to regression R^2 of 50%, Supplementary Datasheet S1). Note that while this analysis did account for all the available covariates, it didn't account for family structure and, thus, had low statistical power.

Discussion

Telomeres are attractive targets for aging, cancer research and reproductive medicine^{6–9}. Since their discovery, a large number of studies have addressed the biology of their inheritance^{27–29} and the causes of inter-individual variation of telomere length, including genetic and environmental factors^{10–14}. These studies have shown that telomere length is a largely heritable trait, and is also affected by both environmental and genetic factors in a population-specific manner^{22,23,26}. Due to involvement of the mentioned factors in telomere length variation, consistency of results across studies strongly depends on the study design, particularly on inclusion of families and multivariate analysis.

Here, we have analyzed whole genome sequencing data obtained by the genome of the Netherlands (GoNL) project, which included 250 Dutch family trios with associated meta-information on sex, age, paternal and maternal ages at conception. We have used computational means for estimation of telomere length and have performed a series of regression and quantitative trait association analyses to reveal inheritance patterns of telomeres and to identify genetic variations affecting their length.

While age plays a role in telomere length variation during a lifetime, our analysis show that its effect is weaker compared to inheritance factors, namely parental telomere lengths, with mother's telomeres being the strongest predictors of offspring MTL. The observed highly heritable nature of telomeres is in agreement with prior studies performed on twins, showing that lymphocyte telomere length at birth largely determines inter-individual variation, while factors regulating telomeres during lifetime are of less importance^{12,41}. However, whether length inheritance occurs stronger through paternal²⁷ or, as in our case, through maternal line²⁸, is still a controversy.

Our results also show positive association of maternal age at conception (MAC) with offspring MTL. Previous studies have largely linked paternal age at conception (PAC) with offspring TL²⁹. This association has been explained by positive selection of sperm cells with longer telomeres with increasing age in males^{27,30,31}. There is limited evidence on MAC effect on offspring TL due to high correlation between biological ages of partners and the difficulty to assess TL dynamics in healthy female oocytes³⁶. Most studies perform regression analysis to support the hypothesis of PAC association with TL and monitor changes in PAC estimates before and after adjustment for MAC⁴⁰. Our observation of positive association of MAC with offspring MTL was also confirmed after partial correlation analysis. It is a matter of future investigations to find out the causes for such a link: whether MAC directly affects TL in oocytes or is indirectly involved in TL regulation in the embryo.

All in all, in the studied population, offspring MTL is positively affected by maternal and paternal MTL and MAC. While these correlations could partially reflect mendelian inheritance of telomeres³², those could also depend on inheritance of genetic factors regulating TL. Identification of telomere-related genetic variants in our cohort was not trivial, due to multiple factors of TL variability that should be accounted for when performing such studies in families. In particular, family association tests implemented in Plink did not allow for inclusion of age as a confounder, while Merlin did not account for heritability of telomeres as quantitative traits. Furthermore, inclusion of genotypes for each SNV into our multiple linear regression model for TL inheritance did not account for family structure. These limitations, along with the small sample size, reduced the statistical power of our tests. Nevertheless, the only consistent hit – *RRM1* – was identified by Plink-based analysis, where age was the missing confounder. This once again confirmed the relatively small influence of age on TL variability among other factors.

RRM1 encodes the catalytic subunit of ribonucleotide reductase (RNR), an enzyme essential for dNTP synthesis and DNA replication and repair processes⁴². In humans, the role of *RRM1* in telomere length regulation is not established yet. However, it is known that the yeast ortholog of this gene, *Rnr1*, is required for telomerase functionality, possibly via facilitation of telomere synthesis via local temporary boosts of dNTPs⁴³. Our data, thus, suggests that *RRM1* may also be involved in telomere length regulation in humans. Importantly, it is known that in humans, *RRM1* is involved in regulation of cellular proliferation, migration and cancer progression, presumably

due to its role in DNA replication⁴⁴. It is also targeted by certain chemotherapeutics or used as a biomarker to predict their effectiveness^{44–46}. Furthermore, polymorphisms and altered expression of this gene have been linked to increased susceptibility to certain cancers^{45,47}. *RRM1*'s role in cancer, however, is only discussed in terms of its involvement in DNA synthesis, and not in terms of telomere length regulation. This study, thus, may serve as a basis for extending the functional studies on *RRM1* to telomeres as well.

Conclusions

All in all, the well-structured and richly annotated dataset provided by the Genome of the Netherlands project has allowed us to reveal several aspects of telomere length inheritance and variability. We have shown that parents' telomere lengths and mother's age at conception play an important role in defining telomere length variability among individuals, along with age and genetic factors. Additionally, our results on association of *RRM1* with MTL and prior studies on its ortholog in yeast, suggest possible implications for this gene in telomere biology in humans.

We also highlight the limitations of currently available computational tools for performing genome-wide association studies on heritable quantitative traits in family trio datasets. Development of these tools and emergence of sequencing datasets on family trios from other populations will further allow for delineating common and population-specific factors involved in telomere length regulation.

Methods

Datasets. We make use of secondary analysis of whole genome sequencing (WGS) data (<http://www.nlgenome.nl/>), European Genome-phenom Archive, EGAS00001000644) previously generated by the Genome of the Netherlands Project (GoNL). The WGS data had been produced on Illumina HiSeq. 2000 platforms with 12x coverage. It consisted of 748 Dutch individuals forming trios of a mother, a father and a child. We also had information about their age and sex, as well as the ages of conception for the mothers and the fathers.

Single nucleotide variation (SNV) data was obtained from the release 5.4 of the GoNL project, where variant calling had been performed on hg19 assembly of the reference genome by the GoNL Consortium³³.

SNV filtering. We have accounted for SNV data passing the filters with VCFtools version 0.1.14, and have removed SNVs with minor allele frequency <5% (70.8% of all SNVs), those with missingness values of <10% (0.3% of all SNVs) and those not passing Hardy-Weinberg equilibrium significance threshold of 0.001 (1.2% of all SNVs) using Plink version 1.9. Overall, genotype information for 5352580 SNVs for 748 individuals were included in the downstream analyses.

Telomere length calculation. We have calculated mean telomere lengths (MTL) from whole genome sequencing data using Computel version 0.4 with its default parameters³⁴. Computel determines MTL by alignment to a special telomeric reference and by comparing the coverage at the telomeric reference to the coverage at the genome reference³⁴. The WGS data contained multiple runs per individual, and we have taken the median MTL per individual across multiple runs.

Regression analysis. Multivariate linear regression analysis was performed to evaluate the correlation between MTL and age and sex in the studied population. Another MLR model was analyzed to study the inheritance of maternal and paternal MTL, as well as the role of MAC and PAC on offspring MTL. Two of the families with missing data for the mother were removed, and two families with discordant age differences at the time of data collection and at conception were also discarded. Overall, MLR were done on 246 families. A set of pairwise regressions on the predictors were performed to estimate dependence between variables, and interaction terms were introduced for correlated predictors. The MLR models were tested by sequential introduction of predictors and interaction terms. The best model was chosen based on maximization of the adjusted R square term: ultimately, from the three best models with similar adjusted R squared values the simplest one was chosen. Partial correlation analyses were performed with R package ppcor and were used to estimate the influence of each predictor when accounting for the rest of the predictors in the model.

Telomere qTL analysis. Association of single nucleotide variations (SNV) with MTL was performed with family based association tests for quantitative traits implemented in Plink 1.9⁴⁸ and Merlin 1.1.2⁴⁹, as well as with a series of MLR models. Both, Merlin and Plink use the between/within model implemented in the QTDT package⁵⁰. However, Merlin uses a maximum likelihood variance component model to correct for family structure, while Plink uses a procedure based on permutations.

We ran Plink with the qfam-total option for identification of quantitative associations with adjustment for sex, and with 100 000 000 permutations to adjust for family structure. One drawback of this approach is that Plink does not handle covariates (other than sex) for family associations, which did not allow to adjust for age. As age was affecting telomere length in our analysis (see Results), we also ran a number of alternative tests.

Separate linear regression tests for parents and children were performed with Plink quantitative association tests, using an additive model for variants and treating age and sex as covariates.

To perform family association tests accounting for age, we used Merlin. For this, we have downloaded genetic maps of 1000 genomes project variants generated by Hapmap 2 project that involved average recombination rates in three different populations (CEU, YRI, and ASN) (from <https://github.com/joepickrell/1000-genomes-genetic-maps>). Only the coding SNVs were included in these tests, and those SNVs with mendelian errors were marked as missing. While Merlin accounted for age, it had the disadvantage of not accounting for quantitative trait inheritance, i.e. inheritance of telomere lengths from parents to offspring, as well as being limited to variants for which recombination rate information is available.

Finally, we also ran multiple linear regression (MLR) analysis by incorporating an additive model of child's genotype for each variant (a variable taking on values from 0 to 2, according to the number of minor alleles in the genotype) into the MLR model of telomere length inheritance obtained from MLR analysis described above. While these regression models do not account for variant linkage and thus had lower statistical power, they do consider telomere length inheritance, age and family structure.

Data availability

This study makes use of the data generated by the Genome of the Netherlands Project. A full list of the investigators is available from www.nlgenome.nl. The datasets analyzed during the current study are available in the European Genome-phenome Archive repository, <https://www.ebi.ac.uk/ega/studies/EGAS00001000644>.

The scripts used for the analyses are available in Github, at https://github.com/lilit-nersisyan/gonl_tql.

Received: 18 July 2019; Accepted: 24 November 2019;

Published online: 10 December 2019

References

- Palm, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* **42**, 301–34 (2008).
- Kim, W. & Shay, J. W. Long-range telomere regulation of gene expression: Telomere looping and telomere position effect over long distances (TPE-OLD). *Differentiation*. **99**, 1–9 (2018).
- de Lange, T. Shelterin-Mediated Telomere Protection. *Annu. Rev. Genet.* **52**, 223–247 (2018).
- Victorelli, S. & Passos, J. F. Telomeres and Cell Senescence - Size Matters Not. *EBioMedicine* **21**, 14–20 (2017).
- Siderakis, M. & Tarsounas, M. Telomere regulation and function during meiosis. *Chromosome Res.* **15**, 667–79 (2007).
- Bojesen, S. E. Telomeres and human health. *J. Intern. Med.* **274**, 399–413 (2013).
- Herrmann, M., Pusceddu, I., März, W. & Herrmann, W. Telomere biology and age-related diseases. *Clin. Chem. Lab. Med.* **56**, 1210–1222 (2018).
- Mons, U. *et al.* Leukocyte Telomere Length and All-Cause, Cardiovascular Disease, and Cancer Mortality: Results From Individual-Participant-Data Meta-Analysis of 2 Large Prospective Cohort Studies. *Am. J. Epidemiol.* **185**, 1317–1326 (2017).
- Maciejowski, J. & de Lange, T. Telomeres in cancer: tumour suppression and genome instability. *Nat. Rev. Mol. Cell Biol.* **18**, 175–186 (2017).
- Rizvi, S., Raza, S. T. & Mahdi, F. Telomere length variations in aging and age-related diseases. *Curr. Aging Sci.* **7**, 161–7 (2014).
- Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.* **45**, 422–427 (2013).
- 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**, 876–82 (1994).
- Andujar, P. *et al.* Smoking, telomere length and lung function decline: a longitudinal population-based study. *Thorax* **73**, 283–285 (2018).
- Patel, C. J., Manrai, A. K., Corona, E. & Kohane, I. S. Systematic correlation of environmental exposure and physiological and self-reported behaviour factors with leukocyte telomere length. *Int. J. Epidemiol.* **46**, 44–56 (2017).
- Daniali, L. *et al.* Telomeres shorten at equivalent rates in somatic tissues of adults. *Nat. Commun.* **4**, 1597 (2013).
- Friedrich, U. *et al.* Telomere length in different tissues of elderly patients. *Mech. Ageing Dev.* **119**, 89–99 (2000).
- Chow, T. T., Zhao, Y., Mak, S. S., Shay, J. W. & Wright, W. E. Early and late steps in telomere overhang processing in normal human cells: the position of the final RNA primer drives telomere shortening. *Genes Dev.* **26**, 1167–78 (2012).
- Martínez, P. & Blasco, M. A. Replicating through telomeres: a means to an end. *Trends Biochem. Sci.* **40**, 504–15 (2015).
- Shay, J. W. & Wright, W. E. Hayflick, his limit, and cellular ageing. *Nat. Rev. Mol. Cell Biol.* **1**, 72–6 (2000).
- Guan, J. Z. *et al.* Change in the telomere length distribution with age in the Japanese population. *Mol. Cell. Biochem.*, <https://doi.org/10.1007/s11010-007-9518-2> (2007).
- Crous-Bou, M. *et al.* Mediterranean diet and telomere length in Nurses' Health study: Population based cohort study. *BMJ*, <https://doi.org/10.1136/bmj.g6674> (2014).
- Zhu, H. *et al.* Leukocyte telomere length in healthy Caucasian and African-American adolescents: relationships with race, sex, adiposity, adipokines, and physical activity. *J. Pediatr.* **158**, 215–20 (2011).
- Hakobyan, A., Nersisyan, L. & Arakelyan, A. Quantitative trait association study for mean telomere length in the South Asian genomes. *Bioinformatics* **32**, 1697–700 (2016).
- Campa, D. *et al.* Genetic determinants of telomere length and risk of pancreatic cancer: A PANDORA study. *Int. J. Cancer* **144**, 1275–1283 (2019).
- Rachakonda, S. *et al.* Telomere length, telomerase reverse transcriptase promoter mutations, and melanoma risk. *Genes Chromosomes Cancer* **57**, 564–572 (2018).
- Crocco, P. *et al.* Population-specific association of genes for telomere-associated proteins with longevity in an Italian population. *Biogerontology* **16**, 353–364 (2015).
- Eisenberg, D. T. A., Hayes, M. G. & Kuzawa, C. W. Delayed paternal age of reproduction in humans is associated with longer telomeres across two generations of descendants. *Proc. Natl. Acad. Sci. USA* **109**, 10251–6 (2012).
- Broer, L. *et al.* Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur. J. Hum. Genet.* **21**, 1163–8 (2013).
- Eisenberg, D. T. A. & Kuzawa, C. W. The paternal age at conception effect on offspring telomere length: mechanistic, comparative and adaptive perspectives. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373**, 20160442 (2018).
- Unryn, B. M., Cook, L. S. & Riabowol, K. T. Paternal age is positively linked to telomere length of children. *Ageing Cell* **4**, 97–101 (2005).
- Kimura, M. *et al.* Offspring's leukocyte telomere length, paternal age, and telomere elongation in sperm. *PLoS Genet.* **4**, e37 (2008).
- Horvath, K. *et al.* Paternal Age and Transgenerational Telomere Length Maintenance: A Simulation Model. *Sci. Rep.* **9**, 20 (2019).
- Genome of the Netherlands Consortium, L. C. *et al.* Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–25 (2014).
- Nersisyan, L. & Arakelyan, A. Computel: Computation of Mean Telomere Length from Whole-Genome Next-Generation Sequencing Data. *PLoS One* **10**, e0125201 (2015).
- Hochstrasser, T., Marksteiner, J. & Humpel, C. Telomere length is age-dependent and reduced in monocytes of Alzheimer patients. *Exp. Gerontol.* **47**, 160–3 (2012).
- Axson, E. L. *et al.* Sex Differences in Telomere Length Are Not Mediated by Sex Steroid Hormones or Body Size in Early Adolescence. *Genet. Genome* **2**, 68–75 (2018).
- Ishida-Takagishi, M. *et al.* The Dishevelled-associating protein Daple controls the non-canonical Wnt/Rac pathway and cell motility. *Nat. Commun.* **3**, 859 (2012).

38. Ara, H. *et al.* Role for Daple in non-canonical Wnt signaling during gastric cancer invasion and metastasis. *Cancer Sci.* **107**, 133–9 (2016).
39. Hoffmeyer, K. *et al.* Wnt/ β -catenin signaling regulates telomerase in stem cells and cancer cells. *Science* **336**, 1549–54 (2012).
40. Eisenberg, D. T. A. & Kuzawa, C. W. The paternal age at conception effect on offspring telomere length: mechanistic, comparative and adaptive perspectives. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373** (2018).
41. Graakjaer, J. *et al.* The relative lengths of individual telomeres are defined in the zygote and strictly maintained during life. *Aging Cell* **3**, 97–102 (2004).
42. Parker, N. J., Begley, C. G. & Fox, R. M. Human gene for the large subunit of ribonucleotide reductase (RRM1): functional analysis of the promoter. *Genomics* **27**, 280–5 (1995).
43. Maicher, A. & Kupiec, M. Rnr1's role in telomere elongation cannot be replaced by Rnr3: a role beyond dNTPs? *Curr. Genet.* **64**, 547–550 (2018).
44. Jordheim, L. P., Sève, P., Trédan, O. & Dumontet, C. The ribonucleotide reductase large subunit (RRM1) as a predictive factor in patients with cancer. *Lancet. Oncol.* **12**, 693–702 (2011).
45. Tóth, C. *et al.* Expression of ERCC1, RRM1, TUBB3 in correlation with apoptosis repressor ARC, DNA mismatch repair proteins and p53 in liver metastasis of colorectal cancer. *Int. J. Mol. Med.* **40**, 1457–1465 (2017).
46. Zhu, C.-M. *et al.* Prognostic value of ribonucleotide reductase subunit M1 (RRM1) in non-small cell lung cancer: A meta-analysis. *Clin. Chim. Acta.* **485**, 67–73 (2018).
47. Xu, X.-L., Zheng, J., Mao, W.-M. & Ling, Z.-Q. RRM1 *151A>&t;T, RRM1 -756T>&t;C, and RRM1 -585T>&t;Gis associated with increased susceptibility of lung cancer in Chinese patients. *Cancer Med.* **5**, 2084–90 (2016).
48. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).
49. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
50. Abecasis, G. R., Cardon, L. R. & Cookson, W. O. A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66**, 279–92 (2000).

Acknowledgements

This study makes use of the data generated by the Genome of the Netherlands project. Funding for that project was provided by the Netherlands Organization for Scientific Research under award number 184021007, dated July 9, 2009 and made available as a Rainbow Project of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). The sequencing was carried out in collaboration with the Beijing Institute for Genomics (BGI). We thank the funders of the sample collections from which the Project Data have been derived: (a) The LifeLines Cohort Study (<http://www.lifelines.nl>), and generation and management of GWAS genotype data for it, is supported by the Netherlands Organization of Scientific Research (NWO, grant 175.010.2007.006), the Dutch government's Economic Structure Enhancing Fund (FES), the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, the University Medical Center Groningen, the University of Groningen, the Dutch Kidney Foundation and Dutch Diabetes Research Foundation; and (b) For sponsorship of the EMC Ergo Study please refer to (<http://www.ergo-onderzoek.nl/wp/>); and (c) The LUMC Longevity Study was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE01014 and IGE05007), the Centre for Medical Systems Biology and the National Institute for Healthy Ageing (Grant 05040202 and 05060810), all in the framework of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.; d. For sponsorship of the VU Netherlands Twin Register please refer to www.tweelingenregister.org. The computational resources for this project were provided by the Academic Scientific Research Computer Network of Armenia.

Author contributions

L.N. has performed experiment design, data analysis, figure and table generation. M.N. has contributed in data analysis and figure, table generation. A.A. has performed experiment design and methodology refinement. All the authors have contributed in manuscript writing/approval.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-55109-7>.

Correspondence and requests for materials should be addressed to L.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Consortia

The Genome of the Netherlands consortium

Laurent C. Francioli³, Androniki Menelaou³, Sara L. Pulit³, Clara C. Elbers³, Wigard P. Kloosterman³, Jessica van Setten³, Isaïc J. Nijman³, Ivo Renkens³, Paul I. W. de Bakker^{3,39}, Freerk van Dijk^{4,5}, Pieter B. T. Neerincx^{4,5}, Patrick Deelen^{4,5}, Alexandros Kanterakis^{4,5}, Martijn Dijkstra^{4,5}, Heorhiy Byelas^{4,5}, K. Joeri van der Velde^{4,5}, Mathieu Platteel⁴, Morris A. Swertz^{4,5}, Cisca Wijmenga^{4,5}, Pier Francesco Palamara⁶, Itsik Pe'er^{6,37}, KaiYe⁷, KaiYe⁸, Eric-Wubbo Lameijer⁸, Matthijs H. Moed⁸, Marian Beekman⁸, Anton J. M. de Craen⁸, H. Eka D. Suchiman⁸, P. Eline Slagboom⁸, Victor Guryev⁹, Abdel Abdellaoui¹⁰, Jouke Jan Hottenga¹⁰, Mathijs Kattenberg¹⁰, Gonneke Willemsen¹⁰, Dorret I. Boomsma¹⁰, Elisabeth M. van Leeuwen¹¹, Lennart C. Karssen¹¹, Najaf Amin¹¹, Fernando Rivadeneira^{11,25}, Aaron Isaacs¹¹, Albert Hofman¹¹, André G. Uitterlinden^{11,25}, Cornelia M. van Duijn¹¹, Mannis van Oven¹², Manfred Kayser¹², Martijn Vermaat^{13,14}, Jeroen F. J. Laros^{13,14}, Johan T. den Dunnen^{13,30}, David van Enckevort¹⁴, Hailiang Mei¹⁴, Mingkun Li¹⁵, Mark Stoneking¹⁵, Barbera D. C. van Schaik¹⁶, Jan Bot¹⁷, Tobias Marschall¹⁸, Alexander Schönhuth¹⁸, Jayne Y. Hehir-Kwa^{19,20}, Robert E. Handsaker^{21,22}, Paz Polak^{21,23}, Mashaal Sohail^{21,23}, Dana Vuzman^{21,23}, Karol Estrada^{21,25,26}, Steven A. McCarrroll^{21,22}, Shamil R. Sunyaev^{21,23}, Fereydoun Hormozdiari²⁴, Vyacheslav Koval²⁵, Carolina Medina-Gomez²⁵, Ben Oostra²⁷, Jan H. Veldink²⁸, Leonard H. van den Berg²⁸, Steven J. Pitts²⁹, Shobha Potluri²⁹, Purnima Sundar²⁹, David R. Cox²⁹, Peter de Knijff³¹, Qibin Li³², Yingrui Li³², Yuanping Du³², Ruoyan Chen³², Hongzhi Cao³², Jun Wang^{32,34,35}, Ning Li³³, Sujie Cao³³, Jasper A. Bovenberg³⁶ & Gert-Jan B. van Ommen³⁸

³Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands. ⁴Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. ⁵Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. ⁶Department of Computer Science, Columbia University, New York, New York, USA. ⁷The Genome Institute, Washington University, St. Louis, Missouri, USA. ⁸Section of Molecular Epidemiology, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands. ⁹European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. ¹⁰Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands. ¹¹Department of Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. ¹²Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. ¹³Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. ¹⁴Netherlands Bioinformatics Center, Nijmegen, The Netherlands. ¹⁵Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ¹⁶Bioinformatics Laboratory, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam, The Netherlands. ¹⁷SURFSara, Science Park, Amsterdam, The Netherlands. ¹⁸Centrum Wiskunde & Informatica, Life Sciences Group, Amsterdam, The Netherlands. ¹⁹Department of Human Genetics, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. ²⁰Center for Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. ²¹Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. ²²Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. ²³Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. ²⁴Department of Genome Sciences, University of Washington, Seattle, Washington, USA. ²⁵Department of Internal Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. ²⁶Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. ²⁷Department of Clinical Genetics, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. ²⁸Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands. ²⁹Rinat-Pfizer, Inc., South San Francisco, California, USA. ³⁰Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands. ³¹Forensic Laboratory for DNA Research, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. ³²BGI-Shenzhen, Shenzhen, China. ³³BGI-Europe, Copenhagen, Denmark. ³⁴Department of Biology, University of Copenhagen, Copenhagen, Denmark. ³⁵The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark. ³⁶Legal Pathways Institute for Health and Bio Law, Aerdenhout, The Netherlands. ³⁷Department of Systems Biology, Columbia University, New York, New York, USA. ³⁸Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. ³⁹Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands.