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

Lilit Nersisyan1*, Maria Nikoghosyan1,2, The Genome of the Netherlands consortium† & Arsen Arakelyan1,2

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 fusions4 and perform a number of regulatory functions, including regulation of gene expression7, 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 cancers6–9.

TL regulation is a complex phenomenon that depends on a number of variables, such as age10, genetic11,12 and environmental13,14 factors, and the tissue under study15,16. 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 replication17,18. 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 some20,21, but not for all populations22,23. 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 populations11,12. Single nucleotide variations (SNVs) in several loci, such as telomerase components TERT and TERC, and others (OBFC1, RTEL1, etc.) have been linked to TL11,24,25. For many loci, however, the identified associations with telomere length are either irreproducible in other cohorts or show...
Population-specifity. Additionally, a number of environmental factors, such as stress, exposure to oxidative agents, smoking, alcohol consumption and lifestyle, all affect telomere attrition rate.

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’s or mother’s TL. Additionally, maternal (PAC) and paternal (MAC) ages at conception also affect TL in the offspring. This has mostly been shown for PAC, and explained by relatively longer TL in the sperm of older males. A recent study has simulated this effect across many generations and has suggested this as a mechanism for adaptation to cancer.

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. 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 age. 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 conception. We have measured mean telomere length (MTL) from WGS data using Computel, 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 consortium. 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 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 homogenous 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 $MTL = Age + Sex$) 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 (Fig. 1). On the other hand, contrary to studies commonly reporting longer telomeres in adult women compared to men, in our cohort sex was not a significant predictor of MTL. Even though females on average had longer telomeres in our population ($5.9 \pm 1.6$ in females versus $5.7 \pm 1.6$ 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 $= 0.442$, Supplementary Fig. S1). There were many variables were not independent, with mother’s and father’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 $MTL = Age + mMTL + fMTL + MAC + PAC$ (adj. $R^2 = 0.442$, Supplementary Table S1).
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, rs1285767 – CCDC88C) are significantly associated with MTL at false discovery rate of 0.1.
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: $\text{MTL} \sim \text{Age} + \text{mMTL} + \text{fMTL} + \text{MAC} + \text{PAC} + \text{SNV}$. 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 $\text{DSCAM}$ 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) ($\text{rs185870422 (chr5q)}$ and $\text{rs1285767 (CCDC88C)}$). The results for these two SNVs and $\text{DSCAM}$ 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, $\text{rs1285767}$, resides within the $\text{CCDC88C}$ gene and leads to 514 bp gain of MTL per additional minor allele. $\text{CCDC88C}$ 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 $\beta$-catenin, which binds transcription factors and regulates expression of $\text{TERT}$, the catalytic subunit of telomerase$^{39}$. 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 1000bp) than homozygotes for this major allele (Fig. 4). This locus distantly interacts with UBLCP1 and RNF145 genes within the same chromatin loop (source: database 3D SNP 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^{-4}-10^{-5}), in contrast to CCDC88C, which had only two SNVs with raw p values < 10^{-5} (Fig. 4, Supplementary Dataset 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 on 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 disease. 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 R² of 50%, Supplementary Dataset 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. Since their discovery, a large number of studies have addressed the biology of their inheritance and the causes of inter-individual variability, possibly via facilitation of telomere synthesis via local temporary boosts of dNTPs. 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 replication. It is also targeted by certain chemotherapeutics or used as a biomarker to predict their effectiveness. Furthermore, polymorphisms and altered expression of this gene have been linked to increased susceptibility to certain cancers. 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.

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 parameters. 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 reference. 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 1.9 and Merlin 1.1.2, as well as with a series of MLR models. Both, Merlin and Plink use the between/within model implemented in the QTDT package. 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-genes-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/lilir-nersisyan/gonl_tql.t.

Received: 18 July 2019; Accepted: 24 November 2019; Published online: 10 December 2019

**References**

1. Pål, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* 42, 301–34 (2008).
2. 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).
3. de Lange, T. Shelterin-Mediated Telomere Protection. *Annu. Rev. Genet.* 52, 223–247 (2018).
4. Vicoirelli, S. & Passos, J. Telomeres and Cell Senescence - Size Matters Not. *ElBioMedicine.* 21, 14–20 (2017).
5. Siderakis, M. & Tarsounas, M. Telomere regulation and function during meiosis. *Chromosoma Res.* 15, 667–79 (2007).
6. Bojesen, S. E. Telomeres and human health. *J. Intern. Med.* 274, 399–413 (2013).
7. Herrmann, M., Pusceddu, I., März, W. & Herrmann, W. Telomere biology and age-related diseases. *Clin. Chem. Lab. Med.* 56, 1210–1222 (2018).
8. Mons, U. et al. 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).
9. Maciejowski, J. & de Lange, T. Telomeres in cancer: tumour suppression and genome instability. *Nat. Rev. Mol. Cell Biol.* 18, 175–186 (2017).
10. Ružič, S., Raza, S. T. & Mahdi, F. Telomere length variations in aging and age-related diseases. *Curr. Aging Sci.* 7, 161–7 (2014).
11. Cord, V. et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.* 45, 422–427 (2013).
12. 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, 876–8 (1994).
13. Andujar, P. et al. Smoking, telomere length and lung function decline: a longitudinal population-based study. *Thorax.* 73, 283–285 (2018).
14. 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).
15. Daniali, L. et al. Telomeres shorten at equivalent rates in somatic tissues of adults. *Nat. Commun.* 4, 1597 (2013).
16. Friedrich, U. et al. Telomere length in different tissues of elderly patients. *Mech. Ageing Dev.* 119, 89–99 (2000).
17. 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).
18. Martinez, P. & Blasco, M. A. Replacating through telomeres: a means to an end. *Trends Biochem. Sci.* 40, 304–15 (2015).
19. Shay, J. W. & Wright, W. E. Hayflick, his limit, and cellular ageing. *Nat. Rev. Mol. Cell Biol.* 1, 72–6 (2000).
20. Guan, J. et al. Change in the telomere length distribution with age in the Japanese population. *Mol. Cell. Biochem.* https://doi.org/10.1007/s10815-007-9518-2 (2007).
21. 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).
22. 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).
23. Hakobyan, A., Nersisyan, L. & Arakelyan, A. Quantitative trait association study for mean telomere length in the South Asian genomes. *Bioinformatics.* 32, 1697–700 (2016).
24. Campa, D. et al. Genetic determinants of telomere length and risk of pancreatic cancer: A PANDoRA study. *Int. J. cancer* 144, 1275–1283 (2019).
25. Rachakonda, S. et al. Telomere length, telomerase reverse transcriptase promoter mutations, and melanoma risk. *Genes Chromosomes Cancer* 57, 564–572 (2018).
26. Crocco, P. et al. Population-specific association of genes for telomere-associated proteins with longevity in an Italian population. *Biogerontology* 16, 353–364 (2015).
27. 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. U.S.A.* 109, 10251–6 (2012).
28. 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).
29. 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, 20160042 (2018).
30. Unruy, B. M., Cook, L. S. & Biabowol, K. T. Paternal age is positively linked to telomere length of children. *Aging Cell* 4, 97–101 (2005).
31. Kimura, M. et al. Offspring's leukocyte telomere length, paternal age, and telomere elongation in sperm. *PLoS Genet.* 4, e37 (2008).
32. Horvath, K. et al. Paternal Age and Transgenerational Telomere Length Maintenance: A Simulation Model. *Sci. Rep.* 9, 20 (2019).
33. Genome of the Netherlands Consortium, L. C. et al. Whole-gene sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* 46, 818–25 (2014).
34. Nersisyan, L. & Arakelyan, A. Computel: Computation of Mean Telomere Length from Whole-Genome Next-Generation Sequencing Data. *PLoS One* 10, 0125201 (2015).
35. Hochstrasser, T., Marksteiner, J. & Humpel, C. Telomere length is age-dependent and reduced in monocytes of Alzheimer patients. *Exp. Gerontol.* 47, 160–3 (2012).
36. Asseo, E. L. et al. Sex Differences in Telomere Length Are Not Mediated by Sex Steroid Hormones or Body Size in Early Adolescence. *Gen. Genome* 2, 68–75 (2018).
37. Ishida-Takagishi, M. et al. The Disevelled-associated 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).
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.emc-ongero-ondernzoek.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. Open access funding provided by Karolinska Institute.

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.

© The Author(s) 2019
Consortia

The Genome of the Netherlands consortium

Laurent C. Francioli, Androniki Menelaou, Sara L. Pulit, Clara C. Elbers, Wiggard P. Kloosterman, Jessica van Setten, Isaac J. Nijman, Ivo Renkens, Paul I. W. de Bakker, Freerk van Dijk, Pieter B. T. Neerincx, Patrick Deelen, Alexandros Kanterakis, Martijn Dijkstra, Heorhiy Byelas, K. Joeri van der Velden, Mathieu Platteeuw, Morris A. Swertz, Cisca Wijmenga, Pier Francesco Palamara, Itsik Pe’er, Kai Ye, Kai Ye, Eric-Wubbo Lameijer, Matthijs H. Moed, Marian Beekman, Anton J. M. de Craen, E. Eka D. Suchiman, P. Eline Slagboom, Victor Guryev, Abdel Abdellau, Jouke Jan Hottenga, Mathijs Kattenberg, Gonneke Willemsen, Dorret I. Boomsma, Elisabeth M. van Leeuwen, Lennart C. Karssen, Najaf Amin, Fernando Rivadeneira, Aaron Isaacs, Albert Hofman, André G. Uitterlinden, Cornelia M. van Duijn, Mannis van Oven, Manfred Kayser, Martijn Vermaat, Jeroen F. J. Laros, Johan T. den Dunnen, David van Enckevort, Hai liang Mei, Mingkun Li, Mark Stoneking, Barbara D. C. van Schaik, Jan Bot, Tobias Marschall, Alexander Schönhuth, Jayne Y. Hehir-Kwa, Robert E. Handsaker, Paz Polak, Mashaal Sohail, Dana Vuzman, Karol Estrada, Steven A. McCarrol, Shamil R. Snyaeve, Fereydoun Hormozdiari, Vyacheslav Koval, Carolina Medina-Gomez, Ben Oostra, Jan H. Veldink, Steven J. Pitts, Shobha Potluri, Purnima Sundar, David R. Cox, Peter de Knijff, Qibin Li, Yingli Li, Yueqing Du, Ruoyan Chen, Hongzh Cao, Jun Wang, Ning Li, Sujie Cao, Jasper A. Boovenberg & Gert-Jan B. van Ommen.

1Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands. 2Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. 3Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. 4Department of Computer Science, Columbia University, New York, New York, USA. 5The Genome Institute, Washington University, St. Louis, Missouri, USA. 6Section of Molecular Epidemiology, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands. 7European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. 8Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands. 9Department of Epigenetics, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. 10Department of Genetic Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. 11Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. 12Netherlands Bioinformatics Center, Nijmegen, The Netherlands. 13Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. 14Bioinformatics Laboratory, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam, The Netherlands. 15SURFsara, Science Park, Amsterdam, The Netherlands. 16Centrum Wiskunde & Informatica, Life Sciences Group, Amsterdam, The Netherlands. 17Department of Human Genetics, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. 18Center for Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. 19Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. 20Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. 21Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts, USA. 22Department of Genome Sciences, University of Washington, Seattle, Washington, USA. 23Department of Internal Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. 24Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. 25Department of Clinical Genetics, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. 26Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands. 27Department of Clinical Genetics, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. 28Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands. 29Department of Clinical Genetics, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. 30Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark. 31BGI-Shenzhen, Shenzhen, China. 32BGI-Europe, Copenhagen, Denmark. 33BGI-Shenzhen, Shenzhen, China. 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, Aardenhout, 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.