Unraveling the causal genes and transcriptomic determinants of human telomere length

Ying Chang[1,#], Yao Zhou[2,#], Junrui Zhou[3,4,#], Wen Li[1], Jiasong Cao[1], Yongmei Shen[1], Xutong Fan[2], Hongxi Yang[2,5], Xiaobao Dong[3], Shijie Zhang[5], Xianfu Yi[2], Ling Shuai[6], Lei Shi[7], Zhe Liu[7], Jie Yang[7], Jihui Hao[8], Kexin Chen[9], Dandan Huang[5,*], Feng Wang[3,*], Mulin Jun Li[2,9,*]

[1]Tianjin Key Lab of Human Development and Reproductive Regulation, Tianjin Central Hospital of Obstetrics and Gynecology, Nankai University, Tianjin, China.
[2]Department of Bioinformatics, The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Key Laboratory of Immune Microenvironment and Disease (Ministry of Education), School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China.
[3]Department of Genetics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China.
[4]Department of Clinical Laboratory, Shanghai Children's Hospital, Shanghai Jiaotong University, Shanghai, China.
[5]Department of Pharmacology, Tianjin Key Laboratory of Inflammation Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China.
[6]State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy, Tianjin Central Hospital of Gynecology Obstetrics/Tianjin Key Laboratory of Human Development and Reproductive Regulation, Nankai University, Tianjin, China.
[7]Key Laboratory of Immune Microenvironment and Disease (Ministry of Education), School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China.
[8]Department of Pancreatic Cancer, Key Laboratory of Cancer Prevention and Therapy, National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China.
[9]Department of Epidemiology and Biostatistics, Tianjin Key Laboratory of Molecular Cancer Epidemiology, National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China.

[#]The authors contributed equally to this work
*Correspondence: mikey.huang2011@gmail.com (D.H.), wangf@tmu.edu.cn (F.W.), mulinli@connect.hku.hk (M.J.L.)

Abstract

Telomere length (TL) shortening is a pivotal indicator of biological aging and is associated with many human diseases. The genetic determinates of human TL have been widely investigated, however, most existing studies were conducted based on adult tissues which are heavily

influenced by lifetime exposure. Based on the analyses of terminal restriction fragment (TRF) length of telomere, individual genotypes, and gene expressions on 166 healthy placental tissues, we systematically interrogated TL-modulated genes and their potential functions. We found that placental TL is relatively longer across human tissues and which maintenance is mostly connected to genes responsible for alternative lengthening of telomeres. Trans-ancestral TL genome-wide association studies (GWASs) on 644,553 individuals identified 20 novel genetic associations and provided increased polygenic determination of human TL. Next, we integrated the powerful TL GWAS with placental expression quantitative trait locus (eQTL) mapping to prioritize 31 likely causal genes, among which 4 were functionally validated, including *MMUT*, *RRM1*, *KIAA1429*, and *YWHAZ*. Finally, modeling transcriptomic signatures and TRF-based TL improved the prediction performance of human TL. This study deepened our understanding of causal genes and transcriptomic determinants of human TL, promoting the mechanistic research on fine-grained TL regulation.

Introduction

Telomeres are DNA and protein complexes that protect the ends of chromosomes, yet degradative processes that shorten telomeric DNA can lead to loss of telomere function and genomic instability [1,2]. Telomeres shorten with each round of DNA replication in the organism's aging process [3]. Thus, telomere length (TL) has been recognized as a critical indicator of cellular senescence, biological aging, and disease progression [4-6]. In previous studies, determinants of human TL have been extensively interrogated, including different genetic, environmental, and lifestyle factors [2,7]. For example, genetic variants associated with TL have been systematically identified through family studies [8,9] and genome-wide association studies (GWASs) [10-12]. Non-genetic factors, such as cigarette smoking [13], alcohol consumption [14], and endurance training [15], could modulate telomere attrition processes. In addition, a large-scale cross-tissue TL analysis revealed that TL varied across tissue types and was the shortest in whole blood [16]. Despite these successes, most of the current studies on human TL were performed on postnatal or adult tissues, which confounds the understanding of independent determinates from genetic or environmental factors, especially for non-Medawarian tissues affected by unobserved lifetime exposures [17,18].

Current TL GWASs have uncovered more than a hundred genomic loci significantly associated with leukocyte TL [10-12], however, the true causal genes underlying these polygenic determinants of TL remain elusive. The majority of the identified associations lie in the non-coding regions of the human genome, suggesting that causal variants could influence TL via gene regulatory codes. Expression quantitative trait loci (eQTLs) analysis using gene expression as a key intermediate molecular phenotype, improves the functional interpretation of GWAS findings [19]. Thus, integrating powerful TL GWAS results with eQTLs on tissue rarely affected by non-genetic factors would enhance the TL-causal gene discovery [20,21] and

also facilitate accurate TL prediction by incorporating both genomic and transcriptomic information.

In the present study, by assuming that telomere is less affected by extraplacental exposure or other non-genetic effects in the placenta, we profiled terminal restriction fragment (TRF) length of telomere, genotypes, and gene expression on 166 healthy placental tissues. We found that placental TL expresses the intra-tissue homogeneity and is longer across human tissues, except for testis. The analysis of gene expression association with placental TL revealed several unique telomere-maintaining patterns. Importantly, we integrated three large-scale TL GWASs from worldwide cohorts and performed a trans-ancestral meta-analysis. Then, placental eQTL mapping and complementary statistical approaches were leveraged to comprehensively prioritize the putative causal genes affecting TL phenotype. We also experimentally validated several novel TL-causal genes *in vitro*. Finally, we developed an accurate TL prediction model that outperforms the existing strategies.

Results

TL in the placenta expresses intra-tissue homogeneity and is longer across tissues

To measure the telomere-associated phenotypes of human tissues at the early life stage, we conducted population-scale southern blots of TRFs [22], an accurate characterization of telomeres, on 166 healthy placental tissues (see methods for details). These placentae were collected within 10 min of a vaginal delivery from full-term singleton pregnancies (37^{+0}–41^{+6} weeks) (Fig. 1A), and none of the participants had any medical disorders or adverse pregnancy outcomes. The average age of the participants was 32 ± 4.0 years, and no tobacco-smoking or alcohol-drinking behavior was noted. Among these newborns, 73 were females and 93 were males. No significant differences were observed in maternal age, maternal body mass index (BMI), infant weight, and gestation weeks with regard to infant gender (all *P*-values > 0.05) (Supplementary Table 1). In addition to these placental tissues, genotypes [the Infinium Asian Screening Array (ASA), n = 166] and gene expression (RNA-seq, n = 166) were also profiled for in-depth analysis of the causal genes and transcriptional determinants of placental TL (Fig. 1B).

Previous studies reported that the TRF length of telomere is highly synchronized among various tissues at birth, whereas in adults, TL across tissue types varies within individuals [16,23,24]. The sampling of eight symmetrical sets of placental tissue on the fetal side and maternal side revealed a consistent distribution of telomere fragments, suggesting a homogeneous telomere status across the whole placenta (Fig. 1C). However, compared to intraplacental telomere status, the ranges and variances in TRF lengths were enlarged as assessed by randomly sampling of 15 unrelated individuals (Supplementary Fig. 1). These results implied that the telomere content of

113 placental tissue varies among individuals and could be attributed to varied genetic conditions
114 and intrauterine environments.

115 Since the southern blot of TRF contains abundant information on telomere content, based on
116 TRF analysis of the fetal side placental tissues from 166 unrelated singleton pregnancies, we
117 quantified the average TRF length (aTL), relative TL (RTL), and short telomere proportion
118 (STP), respectively (see methods for details). Briefly, the RTL was used to evaluate the aTL
119 relative to a standard reference, and the STP measured the percentage of the telomere shorter
120 than 5 kb over fragments, indicating severe telomere damage or wear [25]. Placental aTL ranges
121 from 8.4–17.8 kb among 166 placental tissues (mean = 11.9 kb). Leveraging GTEx RTL data of
122 48 different adult tissues revealed that RTL in the placenta is longer than that of most of other
123 tissues, except the testis (Fig. 1D). This conforms to the expectation that TL is associated with
124 cellular senescence; it is maximal at birth and decreases with age and exposure [26]. No clear
125 association was observed between RTL and the collected demographic factors, including
126 maternal age, gestational days, maternal BMI, infant weight, and placental size (Supplementary
127 Fig. 2). Notably, neonatal sex showed weak evidence of association with RTL in the placenta,
128 wherein males have longer RTL than females (*P*-value = 0.032, *t*-test, Supplementary Fig. 2).
129 Given the limited sample size of our existing TRF measurements in newborns [23,27], the
130 underpowered associations require further ascertainment on large-scale samples. Moreover, the
131 current data revealed that placental STP was negatively correlated with RTL (Fig. 1E),
132 indicating the dependency between TL and short telomere, and RTL could partially explain
133 telomere damage. However, we did not observe significant differences between placental STP
134 and any collected demographic factors (Supplementary Fig. 3). Taken together, the intra-tissue
135 homogeneity of telomere content, long RTL across tissues, as well as less external
136 environmental intervention make placental tissue an ideal proxy for studying the genetic
137 determinants and causal genes of human TL.

138

139 **Trans-ancestral GWAS reveals increased polygenic determination of TL**

140 In order to explore the extent of genetic contribution on TL, we first integrated several
141 large-scale TL GWASs from different cohorts on worldwide populations, including Singapore
142 Chinese Health Study (SCHS) [10], NHLBI Trans Omics for Precision Medicine (TOPMed) [12],
143 and UK Biobank (UKBB) [11]. Next, we conducted a trans-ethnic meta-analysis based on these
144 leukocyte TL GWASs, containing 644,553 participants from five human subpopulations
145 (including European, African, East Asian, South Asian, and Hispanic/Latino) (Fig. 2A, 2B and
146 Supplementary Table 2) (see methods for details). A total of 222 sentinel variants (>10 Mb
147 between sentinels) were associated with leukocyte TL at genome-wide statistical significance
148 threshold (*P*-value < 5E-8), of which 20 variants were new (R^2 < 0.01 with previously
149 documented sentinels) (Supplementary Table 3 and Supplementary Table 4). The
150 meta-analysis revealed 87% (175) of the 201 previously reported TL GWAS significant loci,
151 whereas the remaining 13% (26) loci displayed potential heterogeneity among the three

investigated cohorts. Among the newly discovered loci, the most significant sentinel variant rs10798002 received moderate signals in UKBB and TOPMed cohorts and reached genome-wide significance after meta-analysis, but there was no evidence of the effect of heterogeneity across cohorts (P_{het} = 0.899, I^2 = 0) (Fig. 2C). This variant is located in the *SWT1* gene, affecting the surveillance of nuclear messenger ribonucleoprotein particles [28]. Gene-set enrichment analysis of these TL GWAS signals identified that the most significantly associated pathways were related to telomere maintenance, telomere organization, and telomere maintenance via telomere lengthening (Fig. 2D), which is consistent with the previous findings [11] and implies that TL GWAS signals were related to the regulation of telomere maintenance.

The trans-ancestral TL GWAS meta-analysis, with the largest sample size to date, provided an effective resource to test the agreement between the genetic determination of TL and observed TL, especially that from fetal tissues under minimal extrauterine intervention. Thus, we genotyped and imputed 6,091,762 genetic variants for the 166 placental samples and performed polygenic risk score (PRS) analyses based on the sentinel significant variants of all independent loci from TL meta-analysis results. The estimated PRS score of TL was significantly correlated with placental TL measured in our study (r = 0.21, P-value = 0.007) (Fig. 2E), suppressing the correlations using GWAS summary statistics of the single cohort (Supplementary Fig. 4). This suggested that the trans-ancestral GWAS boosts the predictive power of PRS on TL. Also, no significant differences were observed in the distribution of the individual PRS with respect to sex or maternal age (Supplementary Fig. 5). In contrast, the evaluation of such correlations based on 442 GTEx whole blood samples and 1450 UKBB Chinese leukocyte samples revealed a weak association between TL PRS and tested TL measured by biochemical assays (Fig. 2E). Since GTEx applied a Luminex-based method and UKBB used quantitative polymerase chain reaction (PCR) assay to estimate RTL, these suboptimal measurements may undermine the consistency between PRS-predicted and assayed TL compared to southern blot analysis of TRFs. Collectively, combinatory analysis of trans-ancestral GWAS and placental TL measured by TRF assay indicated that human TL could be determined and predicted only genetically.

Placental TL maintenance is associated with alternative lengthening of telomeres (ALT) and functional gene connectivity

Since 91.5% of sentinel variants of TL GWAS loci are located in the non-coding genomic region, investigating their regulatory potential on gene expression would accurately determine TL. Thus, we profiled transcriptomics using RNA-seq for the 166 placental samples with paired genotypes and TL measurements. The expression levels of major telomerase catalytic subunits, including *TERT, TERC, DKC1, NOP10, NHP2,* and *WRAP53*, were not correlated with placental RTL (Supplementary Fig. 6), whereas *TERC* and *NHP2* expressions were undetectable in placenta (i.e., transcripts per million (TPM) = 0 in all samples). In addition, a

SHELTERIN component *TPP1* showed a moderate correlation with placental RTL (r = 0.19, P-value = 0.015) (Fig. 3A), unlike other protein subunits of SHELTERIN and CTC1-STN1-TEN1 (CST) complexes, such as *TINF2*, *RTEL1*, *POT1*, and *CTC1* (Supplementary Fig. 6). Interestingly, we observed that placental RTL was positively correlated with many components of the ATL pathway, including *ATRX* (r = 0.17, P-value = 0.022), *DAXX* (r = 0.21, P-value = 0.008), and *SMARCAL1* (r = 0.23, P-value = 0.003) (Fig. 3B and Supplementary Fig. 7), implying that TL maintenance in placenta largely relies on ALT mechanism.

Moreover, we found that several canonical telomere maintenance genes were significantly correlated with placental STP, including telomerase catalytic subunits *NOP10* (r = 0.17, P-value = 0.026) and *WRAP53* (r = -0.17, P-value = 0.025), SHELTERIN component *TPP1* (r = -0.18, p = 0.020), and *RTEL1* (r = -0.17, p = 0.031), as well as CST complex member *CTC1* (r = -0.16, p = 0.046) (Fig. 3C and Supplementary Fig. 8). This phenomenon suggested that short telomere phenotype is an effective indicator sensitive to changes in telomere maintenance genes in the placenta. Consistent with the results of placental RTL, placental STP was significantly associated with many genes in the ATL pathway and was negatively correlated with *ATRX* (r = -0.21, P-value = 0.006), *SMARCAL1* (r = -0.22, P-value = 0.005), *RAD52* (r = -0.21, P-value = 0.007), and *POLD3* (r = -0.22, P-value = 0.004) (Fig. 3D and Supplementary Fig. 9), further indicating that ALT pathway is dominant in placental telomere attrition and repair.

To inspect the underlying biological functions associated with placental TL maintenance, we performed two gene module analyses based on transcriptome data from the 166 placentae. First, we used a weighted gene co-expression network analysis (WGCNA) [29] to construct the co-expression networks of the human placenta and identified 26 network modules (Fig. 3E). The correlation test between each module eigengene (ME) score and RTL identified turquoise module was significantly associated with RTL, and the genes in this module were enriched in the ubiquitin-like protein transferase activity (GO:0019787, adjusted P-value = 1.61e-06) (Fig. 3E and Supplementary Fig. 10); this finding was supported by previous studies on the ubiquitin-like proteins on telomere regulation [30,31]. Second, to investigate whether RTL affects gene connectivity in the placenta, we calculated the connectivities of 4679 genes with high variance from the 166 placental RNA-seq data, starting from the upper quarter RTL (left side of Fig. 3F) and subsequently added one sample with shorter RTL and removed one sample with longer RTL to recalculate the connectivities. Based on the hierarchical clustering of the RTL-driven gene connectivities, we detected five gene clusters by Elbow method. For example, 1424 genes in cluster 1 were predominantly interconnected among samples with longer RTL and these genes were significantly enriched in the cell adhesion molecule binding (GO:0050839, adjusted P-value = 1.39e-20), while, 585 genes in cluster 4 were highly linked among samples with shorter RTL and these genes were related to receptor ligand activity (GO:0048018, adjusted P-value = 6.16e-09) (Fig. 3G).

Integrating placental eQTL and TL GWAS for systematic prioritization of TL-causal genes

Recent genetic studies have identified many novel TL-associated genes by eQTL-based methods [11,32], but they mainly employed eQTLs derived from adult tissues, probably leads to biased estimation through unobserved lifetime exposures. A genome-wide *cis*-eQTL mapping on 166 placental samples from Asia population (see methods for details) was conducted to examine novel TL-causal genes based on our trans-ancestral GWAS meta-analysis. Compared to the previous placental eQTLs on European cohort [33], we observed similar distributions of genomic distances to the gene transcription start site (TSS) and end site (TES), but identified 3913 more eQTL-associated genes (eGenes) (Supplementary Fig. 11A). The effect sizes of eQTLs in the two cohorts were also correlated (Supplementary Fig. 11B, r = 0.42, P-value < 2.2E-16). Based on RoadMap chromHMM 15 core states of placenta [34], we observed that most of the placental eQTLs are located in the active chromatin regions and significantly enriched in placental active promoter (TssA) and enhancer (Enh and EnhG) states (Supplementary Fig. 11C-D). These results indicated the validity and the gained power of our placental eQTL mapping.

To systematically prioritize the potential TL-causal genes, we integrated our placental eQTLs and aforementioned trans-ancestral GWAS results using two complementary statistical strategies. First, colocalization analysis via COLOC [35] was performed to test the shared causal variants between gene expression (from eQTL) and TL trait (from GWASs). We identified 53 signals with strong evidence of colocalization between placental eQTL and TL GWAS loci (posterior probability PP4 $\geq$ 0.5, Supplementary Table 5). Second, Fusion Transcriptome-wide association analysis (TWAS) [36] and Summary-based Mendelian Randomization (SMR) [37] were applied to test for a significant genetic correlation between *cis*-expression and GWAS signal. Thus, we observed 64 genes reaching transcriptome-wide significance (false discovery rate (FDR) < 0.1, two-tailed Z-test) in TWAS (Fig. 4A, Supplementary Table 6) and identified 61 genes showing a potential association with TL after HEIDI test (FDR < 0.1) in SMR (Supplementary Table 7), respectively.

The intersection of gene prioritization results from COLOC and the union of TWAS and SMR retrieved 31 likely causal genes related to TL (Supplementary Table 8). These candidates encompassed several genes responsible for canonical telomere regulation, such as telomere length maintenance [38] and telomere end protection [39]. Moreover, some of the mechanisms underlying the causal genes have been explored recently. For example, *TSPYL5* is required to maintain POT1 protein levels and suppresses POT1 poly-ubiquitination and degradation exclusively in ALT cells [40]. *GEN1* is required for telomere replication and prevents the cutting of telomeres [41]. *RFWD3* plays a role in DNA damage response and facilitates translesion DNA synthesis [42]. *ATE1* encodes an arginyltransferase for ubiquitin-dependent degradation and is associated with subtelomeric regulation [43]. We also discovered several novel genes whose TL-related function was rarely documented, such as *RRM1*, *MMUT*, *KIAA1429*, *YWHAZ*,

PEX6, *POLI*, *CDC25B*, and *HDDC2*, and exemplified strong evidence of positive causal associations between genetic determined expressions and TL in genomic loci of four novel genes (Fig. 4B–I). Collectively, the stringent prioritization of TL-causal genes based on placental eQTL and large-scale TL GWAS summary information would provide new insight for understanding TL regulation.

Experimental validations of top prioritized genes in TL regulation

To evaluate the causal effect of the novel hits in our prioritization, we functionally verified the positive regulation of TL by perturbing the four genes screened above: *RRM1*, *MMUT*, *KIAA1429*, and *YWHAZ*. Briefly, HTR8/SVneo cell lines were established by immortalizing a physiological extravillous trophoblast cell by transfection with a plasmid containing the simian virus 40 large T antigen [44]. Stable HTR8/SVneo cell lines were established by the knockdown of the above four genes by shRNA plasmids. Next, we detected whether TL in these cell lines would shorten via southern blots of TRF assays to verify the likely causal correlation between investigated genes and TL.

RRM1 gene encodes the large and catalytic subunit of ribonucleotide reductase (RNR), an enzyme for converting ribonucleotides into deoxyribonucleotides, which is essential for DNA replication and DNA repair processes [45]. Moreover, it has been established that RNR is crucial for telomere elongation to prevent the early onset of replicative senescence in telomerase-negative cells in budding yeast [46]. In addition, a WGS-based telomere length analysis in Dutch family trios [47] showed that *RRM1* might also be involved in telomere length regulation in humans. However, the direct correlation between human RNR and TL has not yet been established. Consequently, the knockdown of *RRM1* gene in eight monoclonal HTR8/SVneo lines consistently showed telomere shortening compared to control cells (Fig. 5A and 5B). One cell clone (shRRM1-2-6) was picked for continuous passage. The results demonstrated that TL was gradually shortened with cell passage (Fig. 5C), further confirming the positive regulatory effect of the *RRM1* gene on TL in the placenta.

MMUT gene encodes the mitochondrial enzyme methylmalonyl-CoA mutase. In humans, the product of this gene is a vitamin B12-dependent enzyme that catalyzes the isomerization of methylmalonyl-CoA to succinyl-CoA. A previous study has shown that vitamin B12 modulates telomere integrity by oxidative stress pathway [48], but the causal correlation between *MMUT* and TL is unknown. We also found that *MMUT* knockdown in most monoclonal HTR8/SVneo lines significantly reduced TL (Fig. 5D and 5E). With continuous passage, the TL of the shMMUT-1-2 monoclonal cell line shortened continually (Fig. 5F). In addition, the inhibition of the other two new genes *KIAA1429* (a vital component of the m6A methyltransferase complex) and *YWHAZ* (tyrosine3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta) showed a similar pattern of telomere shortening (Fig. 5G–J). This functional validation greatly supports the causal association between these enzymes and TL maintenance.

Incorporating genetic and transcriptomic information for accurate TL prediction

This study, together with previous GWAS findings, has strengthened the polygenic basis of TL variation, yet the genetic determinants for TL explained by all genome-wide variants were not substantial (< 10% variance explained) [10-12]. The accurate estimation of TL only relies on a single angle of information, such as genetic or epidemiological factors, which shows a low agreement with actual TL measured by different biochemical assays [16,49]. To improve the performance of TL prediction and facilitate TL-related clinical applications [50,51] when accurate TL measurement (such as TRF-based or fluorescent in situ hybridization (FISH)-based test) is absent, we incorporated transcriptomic information of placental tissue into TRF-based RTL prediction model and systematically compared it to the existing strategies. First, an elastic net regression model was constructed on our 166 placental multi-omics data and individual demographic information. Next, we inspected the ability of TL inference using transcriptional risk score (TRS) [52,53] over static PRS. As a result, the TRS model based on genes regulated by variants linked to TL (r = 0.64, P-value = 1.38e-20, 10-fold cross-validation) outperformed the PRS-based model (r = 0.32, P-value = 3.17e-05) in predicting placental RTL (Fig. 6A and 6B), suggesting that transcriptomic information reflects an additional layer of TL determinants than solely genetic information. Since both PRS- and TRS-based models largely depend on GWAS significant variants, we investigated whether incorporating expression signatures from a specific number of genes could enhance the performance of TL prediction. Thus, the elastic net regression was applied to model the TRF-based RTLs of 166 placentas alone with individual PRS and gender information, resulting in 32 selected genes, independent of PRS, which showed non-zero and significant coefficients (see Methods for details) and found that TL prediction model building on transcriptomic score calculated from the expression of these 32-gene signature (TS-32Gene) suppressed the TRS-based model (r = 0.69, P-value = 3.18e-24) (Fig. 6B). Additionally, the network enrichment analysis by EviNet [54] showed that both signature genes and TL-causal genes were enriched in DNA replication signaling pathway (Supplementary Fig. 12), indicating that TS-32Gene represents a unique and powerful predictor of TL. To further gain predictive performance, we combined transcriptomic signature and genetic determinants of TL on several full models training. Notably, the combination model of TS-32Gene, PRS, and TRS exhibits the best performance (r =0.85, P-value = 1.70e-47) among all trained models (Fig. 6B).

To evaluate the validity of TRF-based RTL prediction model independently, we applied the most practical model (with parsimonious information) only based on TS-32Gene and PRS to GTEx multi-omics data and observed a good agreement between the predicted and observed RTL measured by Luminex-based assay across different GTEx tissues (Fig. 6C, r = 0.26, P-value = 5.92e-45), especially in whole blood, ovary, and esophagus tissues (Supplementary Table 9). This result not only demonstrated the generalizability of our TL prediction model in different contexts but also implied a shared pattern of TL regulation between the placenta and

other tissues. In addition, TL estimated by TelSeq [55], a sequencing-based TL measurement (see Methods for details), showed a weak correlation with Luminex-derived TL in GTEx whole blood tissues (Fig. 6D, r = 0.13, P-value = 0.006), further indicating the superiority of our strategy.

Discussion

Telomere shortening is a classical hallmark of cell senescence and aging, and there are many known elements that contribute to the individual variations of TL, such as genetic, environmental, and lifestyle factors [2]. Although several large-scale GWASs have identified a large number of genetic loci associated with TL [11,12,32]; however, the true causal genes underlying the telomere content regulation are yet to be elucidated. By leveraging TRF assay, genotyping chip, and RNA-seq on hundreds of placenta samples together with trans-ancestral TL GWAS and functional validations, we systematically investigated the causal genes and developed a powerful prediction model for human TL.

A recent study by GTEx consortium has investigated the determinants of TL across various adult human tissues and cell types [16]. However, samples included in GTEx experienced life-course exposures to the external environment and physiological status. The inherently short or long TL might be largely determined at birth and may be crucial for lifelong health [27]. Studies on TL of human tissues with a primitive state are lacking. The placenta embeds in the maternal uterus, allowing nutrition delivery and promoting the growth and development of the fetus [56]. In addition, heritability estimates on adults are often affected by environmental factors, impeding the identification of the true genetic determinants. Since newborn TL could predict later life TL [57], it is critical to investigate TL determinants of the newborn. However, no existing study has harnessed a significant number of early samples without postnatal environmental exposure to study TL.

In the placenta, the telomere is less affected by extraplacental exposure or other non-genetic effects, which provide an ideal proxy for studying the genetic determinants and causal genes of TL. Using the TRF measurement, we are able to not only calculate the mean TL, but also characterize the STP. Next, we found that placental RTL and STP were significantly correlated with genes involved in the ALT pathway but not with telomerase subunit genes. ALT extends the telomeres based on DNA recombination but in a telomerase-independent manner [58]. ATRX interacts with DAXX promoting H3.3 deposition, and participates in DNA replication, genomic stability maintenance, and telomere function maintenance [59]. *ATRX* gene represses ALT, suggesting the role of ALT repression in determining placental RTL. Besides, we observed low expression levels of telomerase genes, including *TERC* and *TERT*. This was consistent with previous findings [60,61] that high ALT activity is associated with low or no telomerase activity. The current results emphasized the importance of genes involved in the ALT pathway in telomere function and length maintenance. Furthermore, we observed a

discrepancy in RTL characteristics between the placenta and other adult human tissues. Consistent with previous studies, the present study revealed that placental RTL exhibits intra-tissue homogeneity but is longer than other somatic adult tissues [23,24,62]. Nevertheless, the placental RTL of male infants is slightly longer than that of female infants, while RTL is longer in females than males in adult tissues, suggesting that the females may sustain long RTL during postnatal developments.

Studies from multiple worldwide cohorts were pooled to perform trans-ancestral GWAS meta-analysis in over 500,000 individuals. The power to detect genome-wide significant signals associated with TL was improved; as a result, we could detect 20 novel genetic associations and recover 87% of the previously reported TL GWAS significant loci. Notably, PRSs of 166 placental samples constructed by TL-associating variants identified via trans-ancestral GWAS were significantly associated with RTL. However, when evaluating PRSs using either GWAS variants from single cohort or genotypes from GTEx whole blood/UKBB leukocyte samples, we only detected weaker RTL correlations, suggesting that leveraging the trans-ancestral GWAS and TRF-based TL measurement could boost the predictive power of PRS on TL. On the other hand, gene expressions in placental tissue are less perturbed through lifetime exposures. Based on genotype and transcriptome profiling from 166 placental samples, we conducted genome-wide *cis*-eQTL mapping and performed TL-causal gene discovery together with our trans-ancestral GWAS results. Complementary statistical methods (such as COLOC, Fusion-TWAS, and SMR) yield 31 likely causal genes related to TL, and some are rarely or inadequately associated with TL, such as *MMUT*, *RRM1*, *KIAA1429*, and *YWHAZ*, which showed a positive regulation of TL. By establishing HTR8/SVneo cell lines with a stable knockdown of these four genes, we validated their functional relevance in maintaining TL via TRF assay. However, the biological mechanisms and tissue/cell-type specificity of these novel TL-causal genes still need an in-depth investigation in the future. Supposedly, some causal genes may not have been detected in this study due to the limitations of sample size, population difference, and tissue specificity of TL maintenance and regulation.

Although TRF assay is a gold standard to quantify TL, it also has many drawbacks; it is time-consuming, less cost-effective, and requires large DNA material [63]. Since the fine-grained determinants of placental TL could be explored effectively using both genetic and transcriptomic information, we constructed several TL prediction models and systematically compared them to the existing strategies. The proposed model relying on transcriptomic signature not only exhibited a great performance in our TRF-based data but also generalized well in the GTEx data. Our model performs better than a WGS-based TL estimation method, TelSeq. The gained performance in the independent datasets, especially in whole blood, lung, and esophagus tissues, highlighted that the regulators of placental TL might have similar biological roles in other tissues. Thus, we speculated that our TL prediction strategy could assist the TL-related clinical applications [50,51] when accurate TL measurement (such as TRF-based or FISH-based test) is not reachable.

425

Methods

Sample collection and processing

The healthy singleton Chinese pregnancies (n = 166) were recruited prior to delivery at Tianjin Central Hospital of Gynecology Obstetrics, China. These study participants did not have any recorded medical disorders or adverse pregnancy outcomes. The hospital ethics committees approved the collection and use of human placental samples (approval no. 2022KY071). All participants provided written informed consent before sample collection. Placentas were treated within 10 min of a vaginal delivery from full-term pregnancies (37^{+0}–41^{+6} weeks). The average age of the participants was 32 ±4.0 years, and no tobacco-smoking or alcohol-drinking behavior was noted. The cohort comprised 73 female and 93 male infants. No significant differences were observed in maternal age, maternal BMI, infant weight, and gestation weeks with regard to infant gender ($P > 0.05$).

Placental full-thickness biopsies of 1.5 × 1.5 cm were collected from the standardized regions at a distance of 2 cm from the umbilical cord. The placenta was placed with the fetal side upwards and orientated with the largest umbilical cord artery on the fetal side of the placenta as a reference location. To avoid contamination by non-target origin cells, the membranes were cut away, and excess blood was removed using sterile filter paper. To analyze the possible intraplacental variation of TL, all the processed biopsies were divided into three parts and sampled at the fetal and maternal layers to obtain eight samples of each placenta, as described by Wyatt et al. [64]. All samples were stored in RNAlater at −80 °C until extraction.

TRF length analysis

TRF method combined with pulsed-field electrophoresis was used to detect the TL length of 166 placental tissues. Genomic DNA isolation kit (Biomiga, BW-GD2211-02) was used to extract genomic DNA from placental tissue and HTR8/SVneo cells in the following experiments. TRF length analysis was applied to measure the TL of these DNA samples. Briefly, 1 μg genomic DNA of each sample was digested with *HinfI* and *RsaI* and then analyzed by agarose gel electrophoresis in 0.5× TBE. Pulsed-field gels [1% (wt/vol)] was run at 6 V, 14 °C for 16 h, and normal electrophoresis gels [0.8% (wt/vol)] were run at 100V, 0–4 °C for 3 h. Subsequent procedures, such as gel depurination, gel denaturation, gel neutralization, DNA transfer to membrane, hybridization with DIG-labeled telomere probe, chemiluminescent detection, and TRF length analysis, were conducted. The sequence of telomere probe is TAACCCTAACCCTAACCCTAACCC. As an internal control, HeLa cell DNA was added in the first lane of each experiment to correct for batch-to-batch variation, and the RTL of 166 placental tissues was estimated by the ratio of their TL *vs*. HeLa DNA's TL. Besides, we also consider the telomere < 5000 bp as the short telomere and the ratio of the band's intensity <

462 5000 bp to the total intensity of the entire telomere band as the STP. The quantification and
463 normalization of TRF length were performed using TeloTool [65].

464

465 **Genotyping, imputation, and quality control**

466 Placental samples were genotyped using the Asian Screening Array (ASA) 750k platform, an
467 Illumina whole-genome single nucleotide polymorphism (SNP) chip designed based on a
468 large-scale East Asian whole-genome sequencing data that encompasses about 750,000
469 markers. Genotype calling by ASA resulted in a dataset of 166 individuals typed at 738,980
470 markers. Data cleaning was performed using PLINK v1.9 [66]. All genotyped variants were
471 subjected to quality control (QC) before imputation. Consequently, variants with (1) call rate (<
472 95%) in all samples, (2) minor allele frequency (MAF) < 0.0001, and (3) departures from
473 Hardy-Weinberg equilibrium (p < 1E-5) were removed. Also, individuals with the following
474 criteria were removed: (1) overall SNP genotyping call rate < 95% and heterozygosity rate > 3
475 SD; (2) genetically inferred sex mismatches between genotype and self-report; (3) related
476 individuals with an identity-by-descent value > 0.1875; Before imputation, we removed SNPs
477 with C/G and A/T alleles to avoid strand flipping. Then, we used Michigan Imputation Server [67]
478 to impute untyped SNPs by borrowing the LD information from all samples using GAsP with
479 Minimac4 for imputation and Eagle v2.4 for phasing. Following imputation, any imputed
480 variant with an imputation quality score < 0.3 or MAF < 0.01 was removed.

481

482 **Transcriptome profiling and quantification**

483 During DNA genotyping and TL measurement, we performed RNA-seq to quantify the
484 genome-wide mRNA expression for 166 placental samples. An equivalent of 3 µg RNA per
485 sample was used as an input material for the RNA sample preparations. RNA-seq was
486 performed using the Illumina NEBNext® Ultra™ RNA sample preparation protocol. The final
487 libraries were sequenced on HiSeq 4000 platform using 150 □bp paired-end chemistry and was
488 run with a coverage goal of 80M reads. Reads containing adapter and ploy-N and those with
489 low quality were removed from raw sequencing reads using fastp [68]. Sequencing QC was used
490 to obtain the overall quality, GC content, and adapter contamination using FastQC
491 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Then, we used STAR v2.5.3a [69]
492 to align the paired-end reads to the reference genome. Gene annotation file was downloaded
493 from GENCODE release 26 (https://www.gencodegenes.org/human/release_26.html).
494 RNA-SeqQC v1.1.9 [70] was applied to count the read numbers mapped to each gene. The genes
495 were selected based on the following expression thresholds: ≥ 0.1 TPM and ≥ 6 reads count in
496 at least 20% of samples.

497

Co-expression network construction

We used WGCNA [29] to construct the co-expression modules and calculate the gene connectivities. The co-expression networks were constructed with the soft power at 9, while other parameters were set at default. The adjacency was transformed into a topological overlap matrix (TOM), and the average linkage hierarchical clustering was performed according to the TOM-based dissimilarity measure. The module eigengene (ME) was the first principal component of a given module and could be considered a representative of the module's gene expression profile. The 26 MEs for the 26 distinct modules were each tested for the correlations with RTL. Then, gene connectivities were determined by calculating the connectivity values using softConnectivity function of WGCNA. Briefly, connectivity describes how strongly a gene is connected to all the other genes in the network. The absolute value of Pearson's correlation coefficient was calculated for all pairwise comparisons of gene expression values across all samples. The Pearson's correlation matrix was then transformed into an adjacency matrix. softConnectivity constructed the adjacency matrix and calculated the connectivity of each gene, i.e., the sum of the adjacency to the other genes. A total of 4679 genes exhibiting the top 30% high expression variance (captures more valid information) were selected for the co-expression analysis. K-means clustering was used to determine the gene groups for the connectivity shift. We also determined the number of clusters by Elbow method. Finally, GO enrichment analysis was carried out via R package clusterProfiler [71].

eQTL mapping

The expression values for each gene were further inverse normal transformed across samples by trimmed mean of M-values [72]. eQTL mapping was performed using tensorQTL[73], a GPU-based method with high efficiency. Next, we used a linear regression model, with top 5 genotype principal components (PCs), age, and 30 PEER factors adjusted. Genotype PCs were computed based on the post-QC genotyping VCF using EIGENSTRAT [74]. To detect *cis*-eQTLs effects, we tested the nominal associations between all variant-gene pairs within a ±1 Mb window around the TSS of each gene and estimated the beta-approximated empirical *P*-values to obtain appropriate significance thresholds based on 10,000 permutations of each gene. Multiple testing corrections were assessed using the Benjamini–Hochberg algorithm, with FDR across all *cis*-eQTL tests within each chromosome estimated. The placental chromatin state regions predicted by chromHMM [34] 15-core state model were downloaded from https://egg2.wustl.edu/roadmap/. We performed Fisher's exact test to investigate whether eQTLs were prone to be located in a specific chromatin state than expected. A two-sided *P*-value and odds ratio were calculated to measure the enrichment of eQTLs in the chromatin state regions.

GWAS meta-analysis

We collected genome-wide summary statistics from three TL trait GWASs based on large-scale individuals, including SCHS [10], NHLBI TOPMed [12], and UK Biobank (UKBB) [11]. First, the variants with MAF < 0.01 were excluded, and then a fixed-effect meta-analysis weighted by sample size of each study was conducted using METAL [75]. Genome-wide statistical significance for the meta-analysis was set at p < 5E-8, HETEROGENEITY mode was set to determine whether the observed effect sizes were heterogeneous across samples. To recognize the novel genetic loci in trans-ancestral meta-analyses, we first identified the associated genetic loci in TOPMed, UKBB, and SCHS meta-analyses at a threshold of p < 5E-8. A locus was defined novel in a trans-ancestral meta-analyses if it did not overlap with any loci of GWAS from a single cohort. Manhattan and Q-Q plots were generated by CMplot [76]. In order to identify the positions of loci containing TL-associated variants, linkage disequilibrium (LD) clumping was conducted using PLINK v1.9. The variants were pruned with the following parameters: a *P*-value cutoff of 5E-8, at a genomic distance of 10 Mb, and R^2 < 0.001 with the lead SNP, using the LD structure of the 1000 Genomes Project as a reference panel.

PRS analysis

Polygenic scores of TL were constructed using PRSice-2 [77] to gauge the associations between reported variations of TL in general populations and in the current study. The scores were computed as the weighted sum of effect allele dosages, as a matrix multiplication of SNP dosages per individual by betas per SNP, i.e., the outcome is a single score of each individual's genetic loading for TL. Our measure of predictive power is the incremental R^2 from adding the score to a regression of the phenotype while adjusting for top five genotyping PCs, sex, and maternal age. The PRS was calculated by summing over all SNPs meeting a set of thresholds, respectively. The null *P*-value of the association of the best-fit GWAS *P*-value threshold was converted to the empirical *P*-value under 10,000 permutations. Pearson's correlations between PRS and RTL were used to compare the PRS analytical performance for the Chinese samples in UKBB, all samples in GTEx, and all samples in this study.

Colocalization analysis

COLOC was applied to colocalize eQTL and TL signals which provided evidence of a putative causal correlation between the eQTL target gene and TL [35]. Herein, we used coloc.abf function implemented in the R package COLOC to perform colocalization analysis. The colocalization with P4 > 0.5 was used as a threshold for colocalized signals. The regional plot was generated

using locuszoom (http://locuscompare.com/) [78], and LD was calculated based on the genotype of all individuals from 1000 Genomes project phase3 [79].

TWAS analysis

The summary-based TWAS was applied for GWAS meta-analysis data using FUSION following the pipeline described on their website (http://gusevlab.org/projects/fusion) [36]. FUSION estimated the heritability of gene expression levels explained by SNPs in *cis* regions to each gene using the mixed-linear model (for instance, BLUP, BSLMM, LASSO, Elastic Net, and Top1 models). The weights for gene expression in the placenta were calculated based on the correlation between SNPs and the placental gene expression while accounting for LD among SNPs. The genes that failed heritability check (heritability *P*-value > 0.01) were excluded from further analyses. We restricted the *cis*-locus to 500 kb on the either side of the gene boundary. Then, the associations between predicted expression of genes and TL were identified by FUSIOIN at default settings. Finally, the proportion of variance in gene expression, *P*-value, and Z-score was obtained from FUSION. TWAS Manhattan plot was generated using TWAS-plotter (https://github.com/opain/TWAS-plotter). TRS was constructed by the genetic value weighted by their Z-score in the TWAS.

SMR analysis

The summary-based SMR method allowed us to infer the causal association between genetically determined gene expression and TL. The SMR test was developed to test the association between the exposure and the outcome using a single genetic variant as the instrumental variable [37]. Based on the assumption of SMR, SNPs are required to affect the TL only through the effects on gene expression. *Cis*-eQTLs were used as the instrumental variables in this analysis. The heterogeneity in dependent instruments (HEIDI) test was carried out to test the existence of linkage in the observed association. Rejecting the null hypothesis (i.e., $P_{HEIDI} < 0.01$) indicated the presence of two or more variants in high LD underlying the association. Thus, we used the default settings in SMR (i.e., MAF > 0.01, excluded SNPs with LD R^2 between top-SNP > 0.90 and < 0.05 and one of each pair of the remaining SNPs with LD R^2 > 0.90), and leveraged the FDR for multiple testing corrections.

Elastic net regression

We also used an elastic net regression model to regress TL on maternal age, infant sex, TL PRS, TRS, and gene expressions. It is a regularized regression method that linearly combines the L1

and L2 penalties of the LASSO and ridge methods [80], emphasizing model sparsity while appropriately balancing the contributions of co-expressed genes. The raw values of all the features are standardized by removing the mean and scaling to the unit variance before training. Optimal regularization parameters were estimated via 10-fold cross-validation. The alpha parameter was set to 0.14, and the lambda value from the best prediction model selected by exhaustive grid search was set to 0.18. The elastic net regression model automatically selected features for building a TL predictor and reported an effect size for each feature. To compare the incremental predictive power of PRS and TRS, we also trained two models that included maternal age, infant sex, and TL PRS only or TRS only. Since our sample size was < 200, we did not leave a testing set for validation but used a 10-fold cross-validation strategy. Instead, GTEx was utilized as an independent validation source, where only GTEx tissues with > 100 samples are used for validation. The correlation (R^2) value between the predicted and the true TL across all samples was used to evaluate the accuracy. Moreover, we compared the performance of the elastic net regression model used in this study and a WGS-based TL estimation tool, TelSeq [55]. The TelSeq tool estimates the average TL using counts of sequencing reads containing a fixed number of telomere signature TTAGGG repeats. A repeat number of 12 and a GC content window of 48–52% was applied in this calculation. TelSeq was also used to estimate the TL for 670 GTEx whole blood samples based on sequence alignment files derived from WGS data, while R^2 between the predicted and the true TL across all samples was used to compare the performance.

Cell culture

HEK 293T cells were grown in DMEM (Corning, USA) supplemented with 10% fetal bovine serum (LONSERA, UY) and 1% penicillin-streptomycin. HTR8/SVneo cells were cultured in 1640 (Corning) medium containing 10% fetal bovine serum and 1% penicillin-streptomycin. The optimal culture conditions in the incubator were 37 °C, 5% CO_2, and humidity of about 95%.

shRNA design and plasmid construction

shRNA sequences were introduced into pLKO.1-puro vectors. The targeting sequences for various shRNAs oligos are as follows:
shMMUT-1: 5'-CCCTTGTATTCCAAGAGAGAT-3';
shRRM1-1: 5''CCCACAACTTTCTAGCTGTTT-3';
shRRM1-2: 5'-GCTGTCTCTAACTTGCACAAA-3';
shKIAA1429-1: 5'-CGGAATATGAAGCAACAAATT-3';
shKIAA1429-2: 5'-CGCTGAGCAAAGTTCTCATAT-3';
shYWHAZ-1: 5'-GCAGAGAGCAAAGTCTTCTAT-3';

shYWHAZ-2: 5'-GCAATTACTGAGAGACAACTT-3';
shUBE2R2-1: 5'-CCAATGTCGATGCTTCAGTTA-3';
shScramble: 5'-CCTAAGGTTAAGTCGCCCTCG-3'.

Establishment of stable cell lines

shRNA plasmids were transfected into HEK 293T cells with polyethylenimine (PEI), according to the manufacturer's instructions. Lentiviral particles produced by HEK 293T cells were released into the DMEM medium. At 48 and 72 h, the lentiviral particle-containing medium was collected and filtered using a 0.45 μm Syringe Filter Unit. HTR8/SVneo cells were cultured in a 6-well plate (400,000 cells/well) for 24 h to achieve 70–80% confluency at the time of infection by 2 mL lentiviral particle-containing medium. After one day post-infection, 2 days of puromycin selection (2 μg/mL), and knockdown determination, single cells were picked and seeded into 96-well plates to generate monoclonal cell lines.

Determination of knockdown by quantitative polymerase chain reaction (qPCR)

When infection and puromycin selection of HTR8/SVneo cells were completed or monoclonal cells from 96-well plates were transferred to 6-well plates, total RNA was extracted using the Eastep® Super Total RNA Extraction Kit (Promega). An equivalent of 1 μγ of RNA was reverse transcribed to synthesize cDNA using HiScript® II Q Select RT SuperMix (Vazyme Biotech). Then, 25 ng of cDNA was used as a template for qPCR analysis with ChamQ Universal SYBR qPCR Master Mix (Vazyme Biotech). The forward and reverse primers for qPCR are listed below:

qMMUT-F: 5'-CAGTTGGAAAAAGAAGACGCTGTA-3';
qMMUT-R: 5'-ATCTGCCTGTTTCGCACTGA-3';
qRRM1-F: 5'-ACTAAGCACCCTGACTATGCTATCC-3';
qRRM1-R: 5'-CTTCCATCACATCACTGAACACTTT-3';
qKIAA1429-F: 5'-GTTGTGCCACCACCAAGAGG-3';
qKIAA1429-R: 5'-AACCCACCACGGGAAGAAAT-3';
qYWHAZ-F: 5'-AGCCATTGCTGAACTTGATACA-3';
qYWHAZ-R: 5'-AATTTTCCCCTCCTTCTCCTG-3';
qGAPDH-F: 5'-TGACAACGAATTTGGCTACA-3';
qGAPDH-R:5'-GTGGTCCAGGGGTCTTACTC-3'.

Data Availability

Processed RNA-seq data, full summary statistics of eQTL and GWAS meta-analysis will be deposited into public repository. The full GWAS summary statistics for UKBB were obtained from https://figshare.com/s/caa99dc0f76d62990195. Individual-level genotype data are available by application to the UKBB (https://www.ukbiobank.ac.uk/register-apply/). The full GWAS summary statistics for TOPMed were applied from the database of Genotypes and Phenotypes (dbGaP), under accession phs001974.v3.p1. The full GWAS summary statistics for SCHS were obtained from https://doi.org/10.6084/m9.figshare.8066999. TRF-based TL measurement of placenta samples are available upon request. TL of various tissues in GTEx were downloaded from (https://www.gtexportal.org/home/datasets). The WGS data of GTEx Whole blood samples were applied from dbGaP, under accession phs000424.v8.p2.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant Numbers 32270717, 32070675, 32170762, 82001579) and the Tianjin Committee of Science and Technology (Grant Numbers 19JCJQJC63600 and 20JCYBJC01400), and Tianjin Key Laboratory of Human Development and Reproductive Regulation (2019XHY06 and 2019XHY07).

Competing interests

The authors declare no conflict of interest.

Contributions

M.J.L., F.W. and Y.C. conceived of the project. Y.Z., J.Z. and D.H. performed the experiments and analyses. W.L., J.C., Y.S., X.F., H.Y., X.D. and X.Y. contributed the data collection and processing. L.S., L.S., Z.L., J.Y., J.H. and K.C. contributed to manuscript polishing and provided analysis suggestions. M.J.L., Y.Z., F.W., Y.C. and J.Z. wrote the manuscript. All authors read and approved the final submission.

Figure Legends

Figure 1. Placental material extraction and measurement. (A) Schematic depicts the collection procedures of placental samples. (B) Workflow of analysis in this study. TRF analysis, genotyping, and RNA-seq were conducted simultaneously for each sample. (C) TL of different sets from the same placental tissue was analyzed by TRF with regular electrophoresis gels. (D) Distribution of RTL across placenta and 24 GTEx tissue types, dashed line indicates the mean RTL in placenta. (E) Scatter plot shows the correlations between short TL and RTL, with a simple linear regression line fitted.

Figure 2. Trans-ancestral TL GWAS and PRS analysis. (A) Manhattan plot of various TL GWASs and GWAS meta-analysis. The x-axis represents the genome in physical order; the y-axis shows -log10 *P*-values for all variants. (B) The quantile-quantile (Q-Q) plot compared the *P*-values generated from this fitted distribution against the observed *P*-values. (C) LocusZoom plot for regional associations of a locus associating with TL in *SWT1* gene, SNPs are colored according to their LD with the lead SNP, rs10798002. The left y-axis shows association *P*-values on the -log10 *P*-values for all SNPs in this locus, the right y-axis shows the recombination rate and the x-axis shows the chromosomal position. The bottom of plot shows the near genes. (D) Dot plot of GO enrichment for the nearest genes of TL-associated variants. (E) Scatter plot for RTL *vs*. PRSs with a simple linear regression line fitted in this study, GTEx, and UKBB.

Figure 3. Gene expression patterns of placental TL maintenance. (A) Scatter plots shows the associations between RTL and *TPP1*. (B) Scatter plots shows the associations between RTL and genes involved in ALT pathway. (C) Scatter plots shows the associations between short TL and *TPP1*. (D) Scatter plots shows the associations between short TL and genes involved in the ALT pathway. (E) Dendrogram of module eigengenes based on dissimilarity measure (1-TOM) and the associations between each module and TL. (F) Heatmap of the connectives for high-variance genes based on about 25% of samples, while the left-most is based on samples with upper-quartile TLs, and the right-most is based on samples with lower-quartile TLs, lower score represents a low overlap and larger score represents a high overlap between the genes. (G) Dot plot of GO enrichment for genes clusters that predominately have higher connectives in samples with long TL (upper) or short TL (bottom).

Figure 4. Causal associations between genetically determined gene expressions and TL. (A) Manhattan plot of transcriptome-wide association results. The x-axis represents the genome in physical order; the y-axis shows Z score for all genes. Genes that passed multiple testing corrections (FDR < 0.1) are highlighted in red and labeled with gene name. (B–E) LocusCompare plots for the (B) *MMUT*, (C) *RRM1*, (D) *KIAA1429*, and (E) *YWHAZ* loci, where the GWAS signals (x-axis) colocalized the eQTL signals (y-axis). LD is colored with respect to the GWAS lead SNPs. (F–I) Scatter plot of the effect sizes of variants reported in TL GWAS and placental eQTLs from (F) *MMUT*, (G) *RRM1*, (H) *KIAA1429*, and (I) *YWHAZ*.

Effect sizes of the variants in the TL GWAS (y-axis) and eQTL (x-axis) are plotted. Error bars indicate 95% confidence interval.

Figure 5. Experimental validation of TL regulation by perturbing expression of *MMUT*, *RRM1, KIAA1429, and YWHAZ.* (A) 34 days after HTR8/SVneo cells were infected with shRRM1 lentiviral particles, 8 shRRM1 monoclonal cell lines grown in 96-well plates were transferred to 6-well plates. TRF assay was used to measure the TL of these cell lines. (B) *RRM1* RNA levels were estimated by qPCR and analyzed by GraphPad Prism software version 6.0. Data are represented as mean ± SD; n = 3. ****P < 0.0001. (C) HTR8/SVneo cells stably expressing control (shScramble) and shRRM1 were passaged over time (DAY) and examined for average TL by TRF. (D) 34 days after HTR8/SVneo cells were infected with shMMUT lentiviral particles, six shMMUT monoclonal cell lines grown in 96-well plates were transferred to 6-well plates. TL of these six cell lines was measured by TRF assay. (E) *MMUT* RNA levels were estimated by qPCR and analyzed by GraphPad Prism software version 6.0. Data are represented as mean ± SD; n=3. ***P < 0.001; ****P < 0.0001. (F) HTR8/SVneo cells stably expressing control (shScramble) and shRNA sequences against MMUT were passaged over time (DAY) and examined for average TL by TRF. (G) 32 days after infection of shKIAA1429 lentiviral particles, six shKIAA1429 monoclonal cell lines from 96-well plates were grown in 6-well plates. The TL of these cell lines was measured by TRF assay. (H) *KIAA1429* RNA levels were tested by qPCR and analyzed by GraphPad Prism software version 6.0. Data are represented as mean ± SD; n = 3. ****P < 0.0001. (I) 32 days after infection of shYWHAZ lentiviral particles, 5 shYWHAZ monoclonal cell lines grew from 96-well plates to 6-well plates. TL of these cell lines was measured by TRF assay. (J) *YWHAZ* RNA levels were estimated by qPCR and analyzed by GraphPad Prism software version 6.0. Data represent mean ± SD; n=3. ****P < 0.0001.

Figure 6. Performance of placental TL prediction model. (A) Scatter plot shows the actual TL values in placenta against the values predicted by the model. (B) Bar chart for performance of elastic net models based on different feature combinations. (C) Scatter plot shows the correlations between TelSeq and TQImean for GTEx whole blood individuals with a simple linear regression line fitted. (D) Scatter plot shows the actual TL values in GTEx tissues against those predicted by the model, only GTEx tissues with >100 samples were used for validation. Dots are colored according to their tissue types.

References

1 McEachern, M. J., Krauskopf, A. & Blackburn, E. H. Telomeres and their control. *Annu Rev Genet* **34**, 331-358, doi:10.1146/annurev.genet.34.1.331 (2000).

2 Blackburn, E. H., Epel, E. S. & Lin, J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science* **350**, 1193-1198, doi:10.1126/science.aab3389 (2015).

3 von Zglinicki, T. Oxidative stress shortens telomeres. *Trends Biochem Sci* **27**, 339-344, doi:10.1016/s0968-0004(02)02110-2 (2002).

4 Fasching, C. L. Telomere length measurement as a clinical biomarker of aging and disease. *Crit Rev Clin Lab Sci* **55**, 443-465, doi:10.1080/10408363.2018.1504274 (2018).

5 Vaiserman, A. & Krasnienkov, D. Telomere Length as a Marker of Biological Age: State-of-the-Art, Open Issues, and Future Perspectives. *Front Genet* **11**, 630186, doi:10.3389/fgene.2020.630186 (2020).

6 Rossiello, F., Jurk, D., Passos, J. F. & d'Adda di Fagagna, F. Telomere dysfunction in ageing and age-related diseases. *Nat Cell Biol* **24**, 135-147, doi:10.1038/s41556-022-00842-x (2022).

7 Andreu-Sanchez, S. *et al.* Genetic, parental and lifestyle factors influence telomere length. *Commun Biol* **5**, 565, doi:10.1038/s42003-022-03521-7 (2022).

8 Njajou, O. T. *et al.* Telomere length is paternally inherited and is associated with parental lifespan. *Proc Natl Acad Sci U S A* **104**, 12135-12139, doi:10.1073/pnas.0702703104 (2007).

9 Slagboom, P. E., Droog, S. & Boomsma, D. I. Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* **55**, 876-882 (1994).

10 Dorajoo, R. *et al.* Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. *Nat Commun* **10**, 2491, doi:10.1038/s41467-019-10443-2 (2019).

11 Codd, V. *et al.* Polygenic basis and biomedical consequences of telomere length variation. *Nat Genet* **53**, 1425-1433, doi:10.1038/s41588-021-00944-6 (2021).

12 Taub, M. A. *et al.* Genetic determinants of telomere length from 109,122 ancestrally diverse whole-genome sequences in TOPMed. *Cell Genom* **2**, 100084, doi:10.1016/j.xgen.2021.100084 (2022).

13 Astuti, Y., Wardhana, A., Watkins, J., Wulaningsih, W. & Network, P. R. Cigarette smoking and telomere length: A systematic review of 84 studies and meta-analysis. *Environ Res* **158**, 480-489, doi:10.1016/j.envres.2017.06.038 (2017).

14 Topiwala, A. *et al.* Alcohol consumption and telomere length: Mendelian randomization clarifies alcohol's effects. *Mol Psychiatry*, doi:10.1038/s41380-022-01690-9 (2022).

15 Werner, C. M. *et al.* Differential effects of endurance, interval, and resistance training on telomerase activity and telomere length in a randomized, controlled study. *Eur Heart J* **40**, 34-46, doi:10.1093/eurheartj/ehy585 (2019).

16 Demanelis, K. *et al.* Determinants of telomere length across human tissues. *Science* **369**, doi:10.1126/science.aaz6876 (2020).

17 Bou Sleiman, M. *et al.* Sex- and age-dependent genetics of longevity in a heterogeneous mouse population. *Science* **377**, eabo3191, doi:10.1126/science.abo3191 (2022).

18 Yamamoto, R. *et al.* Tissue-specific impacts of aging and genetics on gene expression patterns in humans. *Nat Commun* **13**, 5803, doi:10.1038/s41467-022-33509-0 (2022).

19 Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol* **22**, 49, doi:10.1186/s13059-020-02252-4 (2021).

20 Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* **51**, 592-599, doi:10.1038/s41588-019-0385-z (2019).

21 Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet* **11**, 424, doi:10.3389/fgene.2020.00424 (2020).

22 Kimura, M. *et al.* Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nat Protoc* **5**, 1596-1607, doi:10.1038/nprot.2010.124 (2010).

23 Okuda, K. *et al.* Telomere length in the newborn. *Pediatr Res* **52**, 377-381, doi:10.1203/00006450-200209000-00012 (2002).

24 Lin, J. *et al.* Systematic and Cell Type-Specific Telomere Length Changes in Subsets of Lymphocytes. *J Immunol Res* **2016**, 5371050, doi:10.1155/2016/5371050 (2016).

25 Allsopp, R. C. & Harley, C. B. Evidence for a Critical Telomere Length in Senescent Human Fibroblasts. *Experimental Cell Research* **219**, 130-136, doi:https://doi.org/10.1006/excr.1995.1213 (1995).

26 Shammas, M. A. Telomeres, lifestyle, cancer, and aging. *Curr Opin Clin Nutr Metab Care* **14**, 28-34, doi:10.1097/MCO.0b013e32834121b1 (2011).

27 Factor-Litvak, P. *et al.* Leukocyte Telomere Length in Newborns: Implications for the Role of Telomeres in Human Disease. *Pediatrics* **137**, doi:10.1542/peds.2015-3927 (2016).

28 Skruzny, M. *et al.* An endoribonuclease functionally linked to perinuclear mRNP quality control associates with the nuclear pore complexes. *PLoS Biol* **7**, e8, doi:10.1371/journal.pbio.1000008 (2009).

29 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

30 Zalzman, M., Meltzer, W. A., Portney, B. A., Brown, R. A. & Gupta, A. The Role of Ubiquitination and SUMOylation in Telomere Biology. *Curr Issues Mol Biol* **35**, 85-98, doi:10.21775/cimb.035.085 (2020).

31 Grillari, J., Grillari-Voglauer, R. & Jansen-Dürr, P. in *Protein Metabolism and Homeostasis in Aging* (ed Nektarios Tavernarakis) 172-196 (Springer US, 2010).

32 Li, C. *et al.* Genome-wide Association Analysis in Humans Links Nucleotide Metabolism to Leukocyte Telomere Length. *Am J Hum Genet* **106**, 389-404, doi:10.1016/j.ajhg.2020.02.006 (2020).

33 Peng, S. *et al.* Expression quantitative trait loci (eQTLs) in human placentas suggest developmental origins of complex diseases. *Hum Mol Genet* **26**, 3432-3441, doi:10.1093/hmg/ddx265 (2017).

34 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

35 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383, doi:10.1371/journal.pgen.1004383 (2014).

36 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-252, doi:10.1038/ng.3506 (2016).

37 Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-487, doi:10.1038/ng.3538 (2016).

38 Feng, X. *et al.* CTC1-STN1 terminates telomerase while STN1-TEN1 enables C-strand synthesis during telomere replication in colon cancer cells. *Nature Communications* **9**, 2827, doi:10.1038/s41467-018-05154-z (2018).

39 Grandin, N., Damon, C. & Charbonneau, M. Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13. *EMBO J* **20**, 1173-1183, doi:10.1093/emboj/20.5.1173 (2001).

40 Episkopou, H., Diman, A., Claude, E., Viceconte, N. & Decottignies, A. TSPYL5 Depletion Induces Specific Death of ALT Cells through USP7-Dependent Proteasomal Degradation of POT1. *Mol Cell* **75**, 469-482 e466, doi:10.1016/j.molcel.2019.05.027 (2019).

41 Saint-Léger, A. *et al.* The basic N-terminal domain of TRF2 limits recombination endonuclease action at human telomeres. *Cell Cycle* **13**, 2469-2474, doi:10.4161/cc.29422 (2014).

42 Gallina, I. *et al.* The ubiquitin ligase RFWD3 is required for translesion DNA synthesis. *Molecular Cell* **81**, 442-458.e449, doi:10.1016/j.molcel.2020.11.029 (2021).

43 Brower, C. S. *et al.* Liat1, an arginyltransferase-binding protein whose evolution among primates involved changes in the numbers of its 10-residue repeats. *Proceedings of the National Academy of Sciences* **111**, E4936-E4945, doi:10.1073/pnas.1419587111 (2014).

44 Graham, C. H. *et al.* Establishment and characterization of first trimester human trophoblast cells with extended lifespan. *Exp Cell Res* **206**, 204-211, doi:10.1006/excr.1993.1139 (1993).

45 Nordlund, P. & Reichard, P. Ribonucleotide reductases. *Annu Rev Biochem* **75**, 681-706, doi:10.1146/annurev.biochem.75.103004.142443 (2006).

46 Maicher, A. *et al.* Rnr1, but not Rnr3, facilitates the sustained telomerase-dependent elongation of telomeres. *PLoS Genet* **13**, e1007082, doi:10.1371/journal.pgen.1007082 (2017).

47 Nersisyan, L., Nikoghosyan, M., Genome of the Netherlands, c. & Arakelyan, A. WGS-based telomere length analysis in Dutch family trios implicates stronger maternal inheritance and a role for RRM1 gene. *Sci Rep* **9**, 18758, doi:10.1038/s41598-019-55109-7 (2019).

48 Praveen, G., Sivaprasad, M. & Reddy, G. B. Telomere length and vitamin B12. *Vitam Horm* **119**, 299-324, doi:10.1016/bs.vh.2022.01.014 (2022).

49 Puterman, E. *et al.* Lifespan adversity and later adulthood telomere length in the nationally representative US Health and Retirement Study. *Proc Natl Acad Sci U S A* **113**, E6335-E6342, doi:10.1073/pnas.1525602113 (2016).

50 Telomeres Mendelian Randomization, C. *et al.* Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. *JAMA Oncol* **3**, 636-651, doi:10.1001/jamaoncol.2016.5945 (2017).

51 Haycock, P. C. *et al.* Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ* **349**, g4227, doi:10.1136/bmj.g4227 (2014).

52 Marigorta, U. M. *et al.* Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat Genet* **49**, 1517-1521, doi:10.1038/ng.3936 (2017).

53 Liang, Y. *et al.* Polygenic transcriptome risk scores (PTRS) can improve portability of polygenic risk scores across ancestries. *Genome Biol* **23**, 23, doi:10.1186/s13059-021-02591-w (2022).

54 Jeggari, A. *et al.* EviNet: a web platform for network enrichment analysis with flexible definition of gene sets. *Nucleic Acids Res* **46**, W163-W170, doi:10.1093/nar/gky485 (2018).

55 Ding, Z. *et al.* Estimating telomere length from whole genome sequence data. *Nucleic Acids Res* **42**, e75, doi:10.1093/nar/gku181 (2014).

56 Burton, G. J. & Fowden, A. L. The placenta: a multifaceted, transient organ. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140066, doi:10.1098/rstb.2014.0066 (2015).

57 Martens, D. S. *et al.* Newborn telomere length predicts later life telomere length: Tracking telomere length from birth to child- and adulthood. *EBioMedicine* **63**, 103164, doi:10.1016/j.ebiom.2020.103164 (2021).

58 Cesare, A. J. & Reddel, R. R. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet* **11**, 319-330, doi:10.1038/nrg2763 (2010).

59 Valenzuela, M., Amato, R., Sgura, A., Antoccia, A. & Berardinelli, F. The Multiple Facets of ATRX Protein. *Cancers* **13**, doi:10.3390/cancers13092211 (2021).

60 Sung, J. Y., Lim, H. W., Joung, J. G. & Park, W. Y. Pan-Cancer Analysis of Alternative Lengthening of Telomere Activity. *Cancers (Basel)* **12**, doi:10.3390/cancers12082207 (2020).

61 Sung, J. Y. & Cheong, J. H. Alternative lengthening of telomeres is mechanistically linked to potential therapeutic vulnerability in the stem-like subtype of gastric cancer. *Clin Transl Med* **11**, e561, doi:10.1002/ctm2.561 (2021).

62 Allsopp, R., Shimoda, J., Easa, D. & Ward, K. Long Telomeres in the Mature Human Placenta. *Placenta* **28**, 324-327, doi:https://doi.org/10.1016/j.placenta.2006.04.003 (2007).

63 Lai, T. P., Wright, W. E. & Shay, J. W. Comparison of telomere length measurement methods. *Philos T R Soc B* **373**, doi:ARTN 20160451

10.1098/rstb.2016.0451 (2018).

64 Wyatt, S. M. *et al.* The correlation between sampling site and gene expression in the term human placenta. *Placenta* **26**, 372-379, doi:10.1016/j.placenta.2004.07.003 (2005).

65 Gohring, J., Fulcher, N., Jacak, J. & Riha, K. TeloTool: a new tool for telomere length measurement from terminal restriction fragment analysis with improved probe intensity correction. *Nucleic Acids Res* **42**, e21, doi:10.1093/nar/gkt1315 (2014).

66 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

67 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).

68 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890, doi:10.1093/bioinformatics/bty560 (2018).

69 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

70 DeLuca, D. S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530-1532, doi:10.1093/bioinformatics/bts196 (2012).

71 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).

72 Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25, doi:10.1186/gb-2010-11-3-r25 (2010).

73 Taylor-Weiner, A. *et al.* Scaling computational genomics to millions of individuals with GPUs. *Genome Biology* **20**, 228, doi:10.1186/s13059-019-1836-7 (2019).

74 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909, doi:10.1038/ng1847 (2006).

75 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).

76 Yin, L. *et al.* rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Genomics Proteomics Bioinformatics* **19**, 619-628, doi:10.1016/j.gpb.2020.10.007 (2021).

77 Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**, doi:10.1093/gigascience/giz082 (2019).

78 Boughton, A. P. *et al.* LocusZoom.js: Interactive and embeddable visualization of genetic association study results. *Bioinformatics*, doi:10.1093/bioinformatics/btab186 (2021).

79 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

80 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).