GeroScience (2024) 46:5365–5385
https://doi.org/10.1007/s11357-024-01203-2

ORIGINAL ARTICLE

Whole exome sequencing analyses reveal novel genes in telomere length and their biomedical implications

Wei-Shi Liu · Bang-Sheng Wu · Liu Yang · Shi-Dong Chen · Ya-Ru Zhang · Yue-Ting Deng · Xin-Rui Wu · Xiao-Yu He · Jing Yang · Jian-Feng Feng · Wei Cheng · Yu-Ming Xu · Jin-Tai Yu

Received: 29 December 2023 / Accepted: 11 May 2024 / Published online: 5 June 2024

Abstract Telomere length is a putative biomarker of aging and is associated with multiple age-related diseases. There are limited data on the landscape of rare genetic variations in telomere length. Here, we systematically characterize the rare variant associations with leukocyte telomere length (LTL) through exome-wide association study (ExWAS) among 390,231 individuals in the UK Biobank. We identified 18 robust rare-variant genes for LTL, most of which estimated effects on LTL were significant (> 0.2 standard deviation per allele). The biological functions of the rare-variant genes were associated with telomere maintenance and capping and several genes were specifically expressed in the testis. Three novel genes (*ASXL1*, *CFAP58*, and *TET2*) associated with LTL were identified. Phenotypic association analyses indicated significant associations of *ASXL1* and *TET2* with cancers, age-related diseases, blood assays, and cardiovascular traits. Survival analyses suggested that carriers of *ASXL1* or *TET2* variants were at increased risk for cancers; diseases of the circulatory, respiratory, and genitourinary systems; and all-cause and cause-specific deaths. The *CFAP58* carriers were at elevated risk of deaths due to cancers. Collectively, the present whole exome sequencing study provides novel insights into the genetic landscape of LTL, identifying novel genes associated with LTL and their implications on human health

Wei-Shi Liu, Bang-Sheng Wu and Liu Yang contributed equally to the present work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11357-024-01203-2.

W.-S. Liu · B.-S. Wu · L. Yang · S.-D. Chen · Y.-R. Zhang · Y.-T. Deng · X.-R. Wu · X.-Y. He · W. Cheng · J.-T. Yu (✉)
Department of Neurology and National Center for Neurological Diseases, Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Shanghai Medical College, Fudan University, 12Th Wulumuqi Zhong Road, Shanghai 200040, China
e-mail: jintai_yu@fudan.edu.cn

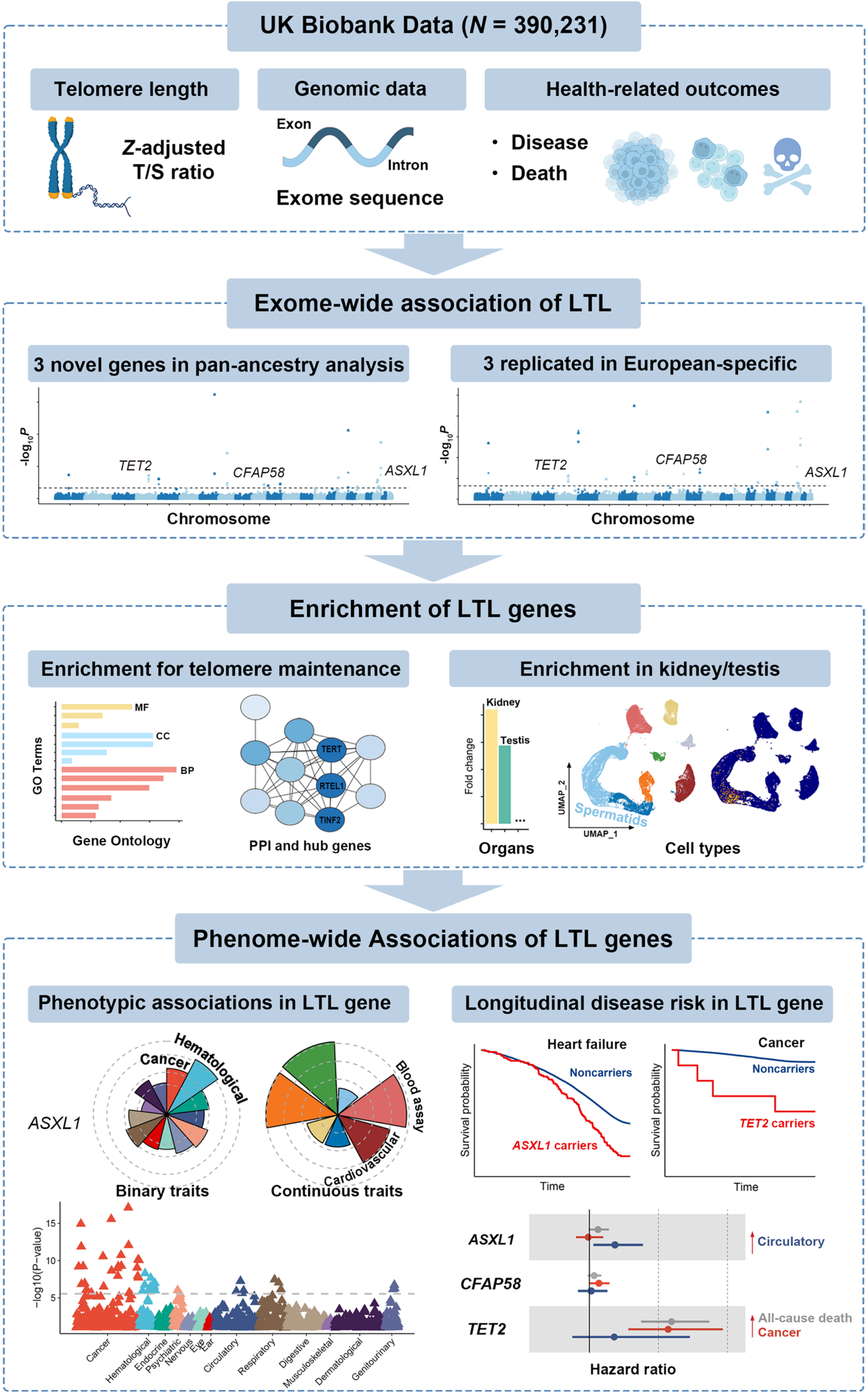
J. Yang · Y.-M. Xu (✉)
Department of Neurology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou University, 1St Eastern Jianshe Road, Zhengzhou 450000, China
e-mail: xuyuming@zzu.edu.cn

J. Yang · Y.-M. Xu
NHC Key Laboratory of Prevention and Treatment of Cerebrovascular Diseases, Zhengzhou, China

J.-F. Feng · W. Cheng
Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

J.-F. Feng · W. Cheng
Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, China

J.-F. Feng · W. Cheng
Department of Computer Science, University of Warwick, Coventry, UK


UK Biobank Data (N = 390,231)
Telomere length
Z-adjusted T/S ratio
Genomic data
Exon
Intron
Exome sequence
Health-related outcomes
• Disease
• Death
Exome-wide association of LTL
3 novel genes in pan-ancestry analysis
3 replicated in European-specific
-log10P
TET2
CFAP58
ASXL1
Chromosome
Enrichment of LTL genes
Enrichment for telomere maintenance
Enrichment in kidney/testis
GO Terms
MF
CC
BP
Gene Ontology
TERT
RTEL1
TINF2
PPI and hub genes
Kidney
Testis
Fold change
...
Organs
UMAP_2
UMAP_1
Spermatids
Cell types
Phenome-wide Associations of LTL genes
Phenotypic associations in LTL gene
Longitudinal disease risk in LTL gene
ASXL1
Cancer
Hematological
Blood assay
Cardiovascular
Binary traits
Continuous traits
−log10(P−value)
15
10
5
Cancer
Hematological
Endocrine
Psychiatric
Nervous
Eye
Ear
Circulatory
Respiratory
Digestive
Musculoskeletal
Dermatological
Genitourinary
Heart failure
Noncarriers
ASXL1 carriers
Survival probability
Time
Cancer
Noncarriers
TET2 carriers
Survival probability
Time
ASXL1
CFAP58
TET2
Circulatory
All-cause death
Cancer
Hazard ratio

◀**Fig. 1** Study design. Top part, UK Biobank data used in the present study, including leukocyte telomere length (LTL), exome sequence data, and health-related outcomes. Middle-upper part, exome-wide association analysis of LTL revealing novel rare-variant genes associated with LTL. Middle-lower part, biological functions of the rare-variant genes, including Gene Ontology (GO) analysis and protein–protein interaction (PPI) network, and tissue expression enrichment analysis. Bottom part, phenotypic association analyses of LTL genes with diseases and biomedical phenotypes. Survival analyses of LTL genes with incident diseases and mortality. BP, biological process; CC, cellular component; MF, molecular function

and facilitating a better understanding of aging, thus pinpointing the genetic relevance of LTL with clonal hematopoiesis, biomedical traits, and health-related outcomes.

Keywords Telomere length · Whole exome sequencing · Aging · Human health

Introduction

Telomeres are the genomic complexes at the end of eukaryotic chromosomes that shorten with each round of cell division and are involved in the maintenance of genomic and cellular stability [1, 2]. With a better understanding of the multiple molecular mechanisms of aging, telomeres are considered an instigator or amplifier of the molecular circuits that drive the aging process and age-related diseases [3]. Telomere shortening, as one of the hallmarks of aging [4], is commonly observed during normal aging in humans [5], and mounting evidence has suggested that shorter telomere length (TL) is associated with the risk of malignant neoplasms, coronary artery disease (CAD), pulmonary fibrosis, and multiple other age-related diseases [6–8]. Therefore, identifying the determinants of telomere shortening would advance a better understanding of the initiating pathophysiology in aging and age-related diseases. Telomere length is inheritable, as previous studies have demonstrated that TL was a highly polygenic trait with a high heritability [9] and revealed numerous genes associated with TL through genome-wide association study (GWAS) and whole-genome sequencing (WGS) study [10, 11]. However, most previous studies have focused on the common genetic variants, with only a few profiling the contribution of rare variants to the variations of TL [12]. In addition, though considerable loci or genes associated with TL have been identified, their estimated effect sizes were generally modest [10]. Exome sequencing mainly focuses on protein-coding variants, particularly rare and ultra-rare variants that have not been genotyped, and directly implicates genes in phenotype variability through the burden testing of multiple rare variants [13]. The growing exome sequencing data and the development of statistical association analyses for the rare variants made it possible to explain the missing heritability, which cannot be interpreted by common variants alone [14], thus providing novel insights into the genetic architecture of TL [13, 15]. In addition, sequencing studies and subsequent gene therapeutic interventions have demonstrated the great translational potential of the rare variants [16, 17]. Therefore, revealing the genetic determinants of TL, particularly the protein-coding variants, will provide insights into the regulation of TL shortening or variation, and identify underlying therapeutic targets for aging and age-related diseases.

Here, we leverage the whole exome sequencing (WES) data in 390,231 participants from the UK Biobank to perform rare variants and gene burden exome-wide association study (ExWAS) of leukocyte TL (LTL) measurements. Eighteen robust rare-variant genes associated with LTL were identified, three of which have not been reported in previous studies. Next, we characterize the biological functions, protein–protein interaction network, and specific tissue expression of the identified genes. We then explore the phenotypic association of the significant rare-variant genes across multiple biomedical phenotypes. Finally, we test the longitudinal risk of age-related diseases and overall and disease-specific mortality among the novel LTL rare-variant gene carriers (Fig. 1).

Methods

Participants of the study

The UK Biobank is a large-scale population-based prospective cohort study in the UK from March 2006 to December 2010, which recruited over 500,000 participants aged 40 to 69 years old at baseline [18]. At recruitment, the demographic, clinical, and genetic information of the participants were collected through touch-screen questionnaire, computer-assisted interview, physical measurements, and biological sample assays, as described in detail elsewhere [18]. All participants have given written informed consent for genomic data and medical records at recruitment.

Measurement of LTL

The measurement, extensive quality checks, adjustment, and transformation of LTL of the participants in UKB have been described in detail elsewhere [19]. Briefly, LTL was measured on DNA extract from peripheral blood samples collected at recruitment using quantitative polymerase chain reaction (qPCR) method, and reported as a ratio of the telomere repeat number to a single-copy gene (T/S ratio). Then, the measurement of LTL was $\log_e$-transformed to approximate the normal distribution and then *z*-standardized to facilitate the comparisons with other datasets [19]. Regarding the further analyses in the present study, participants with LTL measurement at baseline were included ($n = 472,174$).

Since LTL is influenced by multiple conditions, including multi-morbidities, modifiable factors, and behaviors [20], the participants with extreme values (top and bottom 0.5%) were excluded from the analyses to avoid the influence of potential confounding factors.

Exome sequencing and quality control

Whole exome sequencing was conducted on 454,787 participants from the UK Biobank [13]. Our ExWAS analysis was restricted to White British individuals with available LTL and exome sequencing data, further carrying out other quality controls. The IDT xGen Exome Research Panel version 1.0 including supplemental probes on the NovaSeq6000 platform was used to capture the exomes and the full sequencing protocols were described in detail elsewhere [21]. Initial quality check included sex discordance, contamination, unresolved duplicate sequences, and discordance with microarray genotyping data checks. In addition, we performed additional quality control and filtering. We applied genotype refinement to the raw genotype calls in the pVCF files using Hail. We first split multi-allelic sites to represent separate bi-allelic sites. All calls that did not pass the following hard filters were then set to no-call in our analysis. For homozygous reference calls, we chose Genotype Quality < 20, Genotype Depth < 10, and Genotype Depth > 200. For heterozygous calls, (A1 Depth + A2 Depth) / Total Depth < 0.9, A2 Depth / Total Depth < 0.2, Genotype likelihood [ref / ref] < 20, Genotype Depth < 10, and Genotype Depth > 200. For homozygous alternative calls, (A1 Depth + A2 Depth) / Total Depth < 0.9, A2 Depth / Total Depth < 0.9, Genotype likelihood [ref / ref] < 20, Genotype Depth < 10, and Genotype Depth > 200 were used in this work. Regarding the variants, we excluded the variants with low genotype quality, extremely low or high genotype depth, call rate less than 90%, and Hardy–Weinberg *p*-value less than 10^{-15}. For samples, we excluded those with low exome sequencing quality, withdrawn from the study, with duplicates, with discordance between self-reported and genetically inferred sex, and whose call rates or additional metrics were outliers. Moreover, the related participants had high level of genetic correlation, which probably influence the robustness of exome-wide association analysis. To improve the reliability and accuracy of the results of exome-wide association analysis, the participants related at 3rd degree or closer were also excluded. Overall, 390,231 individuals were included in the ExWAS analysis of which 327,790 of them were European. The overall workflow of the quality control of exome sequencing data could be found in Fig. S1.

Exome-wide association analysis

For rare variants, the SKAT-O test through the SAIGE-GENE + method was used to analyze the 20,103 genes in the exome sequencing data [22]. The SAIGE-GENE + method is used for region-based association analysis that is capable of processing large-scale samples, particularly biobank [22, 23]. SAIGE-GENE + can collapse the ultra-rare variants (which are defined as minor allele carrier (MAC) ≤ 10) to a single marker and then test the collapsed variant together with all other variants with MAC > 10, thereby reducing the data sparsity due to the effects of ultra-rare variants [22]. Regarding rare variants, here we applied five different maximum minor allele frequency (MAF) cutoffs (1%, 0.1%, 0.01%, 0.001%, and 0.0005%) and two different variant annotations (LOF and missense), followed by aggregating multiple SKAT-O tests using the Cauchy combination or minimum p-value for each gene or region [24, 25]. SnpEff Version 5.1 was used to annotate and classify the variants of all samples [26]. The LOF variants included the variants annotated as frameshift, splicing donor, splicing acceptor, and stop gain. The missense variants include the variants predicted as deleteriousness in Sorting Intolerant From Tolerant (SIFT) [27], Polymorphism Phenotyping v2 (PolyPhen2) HDIV [28], and PolyPhen2 HVAR [28]; likelihood ratio test (LRT) [29]; and MutationTaster [30], and are further collapsed for each gene. The ExWAS model was adjusted by age at the recruitment, gender, and first ten PCs.

In single rare variant analysis, the variants at exome-wide significance ($p < 1 \times 10^{-8}$) were considered significant. In rare-variant gene–based analysis, Bonferroni correction was used for multiple comparisons (adjusted p-value = 0.05/X/Y/Z, where X represented the number of genes, Y represented the number of maximum MAF cutoffs, and Z represented the number of variant annotations).

Phenome-wide association analysis

For rare-variant genes for LTL, we performed phenome-wide association analysis between the rare-variant genes significantly associated with LTL and 17,361 binary and 1419 continuous phenotypes in the 394,695 European UKB individuals using AstraZeneca PheWAS (https://azphewas.com) [31]. Any phenotype with a p-value < 0.05 was considered nominally significant phenotypes. We then categorized the binary and continuous phenotypes and calculated the proportion of the significant phenotypes in each category. To identify the significant associations, we focused on rare protein-truncating variant (PTV, which was defined as any variant that is predicted to truncate the protein) and Bonferroni corrections were used for multiple comparisons (for binary traits, adjusted $p = 0.05/17361 = 2.88\text{E}-6$; for continuous traits, adjusted $p = 3.52\text{E}-5$).

Gene Ontology analysis

Gene Ontology (GO) analysis was conducted using the R package clusterProfiler [32]. All genes listed in the database were used as background. And the GO terms of three ontologies (biological process, molecular function, and cellular component) with an adjusted $p < 0.05$ (calculated by the method BH) were defined as an enrichment of the GO term. The results of GO analysis were visualized by the R package ggpubr (https://github.com/kassambara/ggpubr).

Protein–protein interaction analysis

The online website STRING (https://string-db.org; version 11) was used to perform protein–protein interaction (PPI) analysis with default parameters. Cytoscape software (version 3.9.1) [33] was used to visualize the result of PPI analysis and hub genes within the network were identified using the cytoHubba plug-in [33].

Tissue enrichment of LTL genes

Tissue enrichment analysis of the genes significantly associated with LTL was performed through the R package TissueEnrich [34] to identify whether the specific gene was enriched in multiple tissues [34]. TissueEnrich uses hypergeometric test to calculate the enrichment of the tissue-specific genes, and the RNA sequencing data from Genotyped Tissue Expression (GTEx) v8 was used [34]. The genes with expression level greater than 1 (TPM) that also have at least five-fold higher expression levels in a certain tissue compared to all other tissues were considered tissue-enriched [35].

Single-cell RNA sequencing data

Single-cell RNA sequencing (scRNA-seq) dataset of healthy human testis was acquired from Gene Expression Omnibus (GEO) with the accession number GSE182786 [36]. We used the R package Seurat to process the scRNA-seq data [37]. First, the individual cells with low quality, which were defined as the cells with less than 800 expressed genes or larger than 50% mitochondrial counts, were excluded and then the gene expression matrix was normalized by the NormalizeData function in Seurat [36, 37]. We used the R package Harmony [38] to integrate the multiple datasets to correct the batch effect. The top 40 PCs and a resolution of 0.1 were used to conduct clustering, and then, the clusters were annotated according to the known markers of each cell type (spermatid: *PRM3*, *SPATA3*; Leydig cell: *IGF1*, *CFD*; endothelial cell: *VWF*, *CD34*; spermatocyte: *SYCP1*, *SYCP3*;

smooth muscle cell: *CRIP1*, *MCAM*; spermatogonia: *MAGEA4*, *ID4*; Sertoli cell: *SOX9*, *DEFB119*; macrophage: *CD14*, *CD163*, *TYROBP*) [36, 37].

Longitudinal survival analyses of rare-variant genes

Longitudinal survival analyses were performed by Cox proportional hazard models using the R package survival (https://github.com/therneau/survival). The carriers of the LTL rare variant were defined as those carrying genetic variant (including both LOF and missense variants) within the gene region, and we have confirmed that all the LTL rare-variant carriers only had 1 of the pre-identified variants in the genes. The primary outcomes included cancers, hematological and circulatory diseases, which were extracted from the UK Biobank health outcome datasets first occurrences of health outcomes (Category 1712, https://biobank.ndph.ox.ac.uk/ukb/label.cgi?id=1712), and all-cause or disease-specific death, which were extracted from the UK Biobank health outcome datasets underlying (primary) cause of death (Field 40,001) based on International Classification of Disease-10. Regarding the diseases, the end of follow-up was defined as the date of the diagnosis of the disease or the end of hospital inpatient data collection. And regarding the death, the end of follow-up was defined as the date of death or the end of hospital inpatient data collection. In each model, the individuals with any diagnosis of the disease or cancer prior to the time at recruitment were excluded. Assumptions of proportional hazards were tested based on Schoenfeld residuals [39]. The hazard ratios (HRs) were adjusted for age at recruitment, gender, and top 10 PCs.

Genomic association analysis

The imputed genotypes available in the UK Biobank v3 imputed genetic data were used for GWAS and only White British participants were included in the analysis [40]. The participants with a missing genotype rate more than 0.05, a mismatch between self-reported (Field 31) and genetic gender (Field 22,001), abnormal sex chromosomal aneuploidy, heterozygosity rate outliers, and exceeding 10 putative third-degree relatives were filtered out. For quality control, we excluded the variants with MAF $<$ 0.01, call rate $<$ 0.95, and imputation quality score $<$ 0.5, falling the Hardy–Weinberg equilibrium test at p-value $< 1 \times 10^{-6}$, or duplicated [41]. In addition, we further excluded the multi-allelic variants in the analysis. Then, PLINK 1.9 was used to perform the association analysis for LTL [42]. Age, gender, and the first ten PCs were considered the covariates. In GWAS, the SNPs at genome-wide significance ($p < 5 \times 10^{-8}$) were considered significant associations.

Mendelian randomization

We first screened all 42,335 traits available in the MR-Base platform [43]. Then, to increase statistical power and avoid significant sample overlap, we excluded the traits with non-European ancestry, small sample size (≤ 50,000), only one gender, or generated from the UK Biobank. In addition, we also excluded the duplicated traits (through systemically considering the sample size and throughput of sequencing). After the screening, 186 traits remained and were further divided into 126 exposures and 61 outcomes. Regarding the selection of the instrumental variables, the single-nucleotide polymorphisms (SNPs) associated with the traits or LTL at genome-wide significance ($p < 5 \times 10^{-8}$) were first screened and then clumped at $R^2 < 0.001$ at a 10,000 kb window size based on the 1000 Genomes European reference panel using PLINK v2.0 [42]. We used the R package TwoSampleMR [43] to perform MR analysis and the inverse-variance weighted method was mainly used to estimate the causal effect. Moreover, MR-Egger, weighted median, weighted mode, and simple mode methods were also performed as additional analysis [43–45]. A p-value less than 0.05 was considered a nominally significant association.

Results

Exome-wide rare variant analysis of LTL

In the analysis of exome sequencing data, a total of 390,231 participants with both LTL and exome sequencing data from the UK Biobank were included after quality controls of genotype, variant, and sample (Fig. S1). We first conducted ExWAS of LTL for the rare variants and found 88 single rare variants (MAF $<$ 0.01) in *CFAP58*, *CTC1*, *DCLRE1B*, *EXOC3L1*, *HBB*, *PARN*, *POT1*, *RTEL1*, *SAMHD1*, *TERF1*, *TERT*, and *TET2*, associated with LTL at exome-wide significance ($p < 1 \times 10^{-8}$; Table S1). Next, we performed whole exome gene–based collapsing tests of LTL (Fig. 2A, B) and identified 19 significant (Bonferroni-adjusted p-value = 0.05/20103 /5/2 = 2.49×10^{-7}) rare-variant genes associated with LTL (Table 1). The *Q-Q* plots for the rare variants are available in Figs. S2 and S3.

To test the robustness of the rare variants and genes identified above, we conducted ExWAS of LTL for the rare variants in the European population. We found that 71 of 88 single rare variants were still associated with LTL at exome-wide significance ($p < 1 \times 10^{-8}$; Table S1). In addition, 18 of 19 rare-variant genes were also significant in the European population after multiple comparisons, except for *HBB* (Table 1). The *Q-Q* plots for the rare variants of European ancestry are available in Figs. S4 and S5.

Of the 18 robust rare-variant genes, thirteen of them have been reported in previous GWAS or WGS studies of LTL (Fig. S6) [10, 11], while five rare-variant genes (*ASXL1*, *CFAP58*, *TET2*, *ZNF451*, and *ZSWIM3*) were not reported in previous GWAS or WGS studies (Fig. S6). Moreover, *ZNF451* and *ZSWIM3* were identified as underlying genes for LTL by Kessler et al. [46]. Among the three novel genes identified from the ExWAS collapsing test, the burden of rare LOF variants in *TET2* showed the most significant associations with LTL ($p = 5.19 \times 10^{-15}$).

Landscape of novel rare-variant genes with LTL

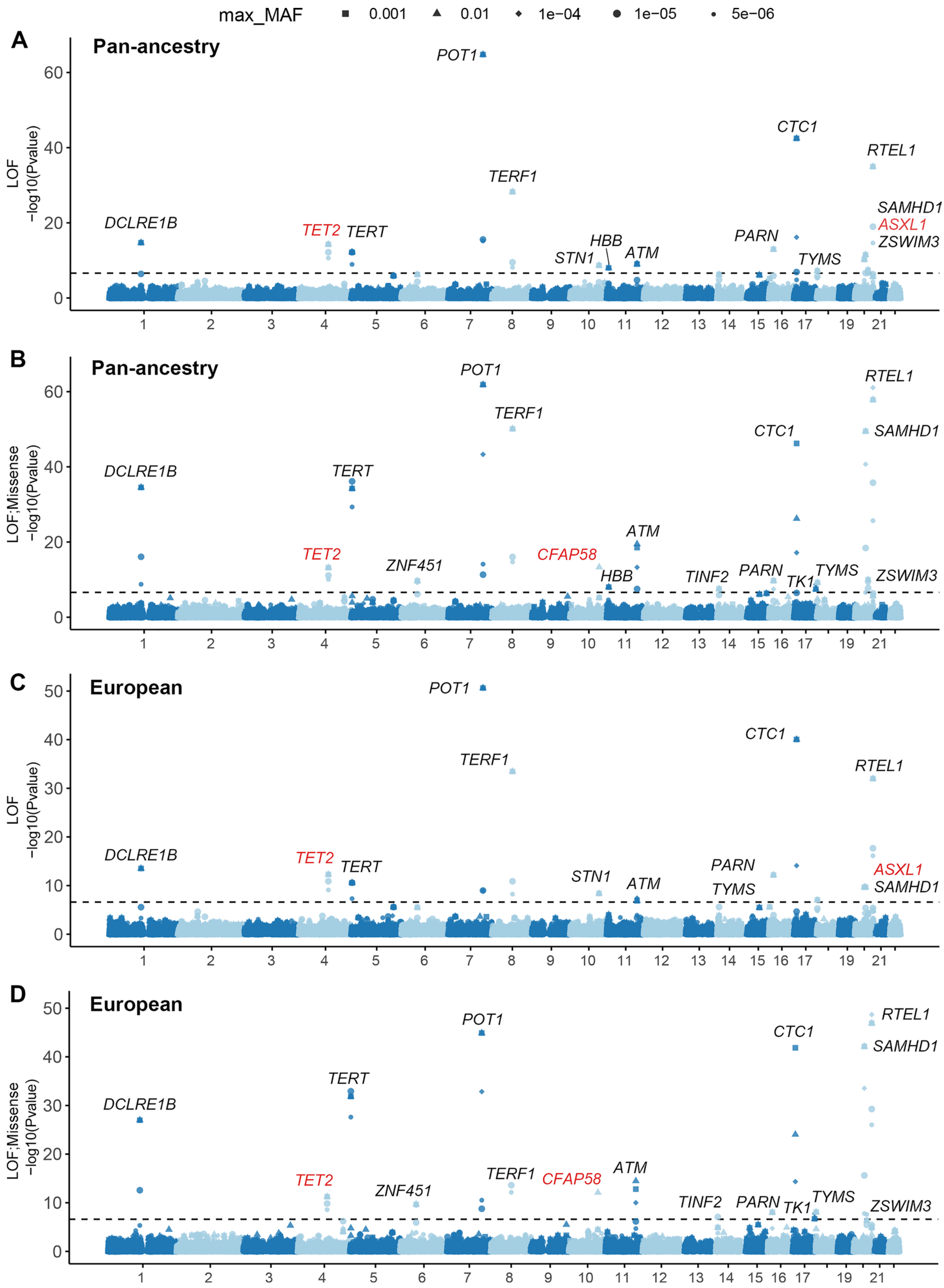
We profiled the carriers with rare-variant genes for LTL of all participants in the UK Biobank (Fig. 3A). Among the novel genes identified associated with LTL, the proportion of participants with *CFAP58* (1.69%, 847 LOF and 6479 missense) variants or *ASXL1* (0.68%, 241 LOF and 2690 missense) was relatively high. In contrast, the number of participants with *TET2* (0.11%, 460 LOF and 11 missense) or *ZSWIM3* (0.11%, 52 LOF and 416 missense) variants was relatively low. Measurements of LTL among the novel LTL gene carriers and noncarriers were also profiled (Fig. 3B). The participants with *ASXL1* (-0.055 ± 0.963 [mean ± standard deviation, *Z*-adjusted T/S log relative LTL] vs. -0.004 ± 0.943, $p < 0.01$) or *TET2* (-0.415 ± 0.962 vs. -0.004 ± 0.943, $p < 0.001$) variants showed shorter LTL. The trend was more significant when we restricted the participants with LOF variants for *ASXL1* (-0.627 ± 0.901 vs. -0.004 ± 0.943, $p < 0.001$) (Fig. 3B). In addition, participants with *CFAP58*, *ZNF451*, or *ZSWIM3* variants had longer LTL than noncarriers.

We also profiled the effects of the genes on LTL via a linear regression model. After adjustment for age at recruitment, gender, and the top 10 PCs, the estimated effects of LTL gene carrier ranged from −0.84 to 0.69 (Fig. 3C). Similarly, the effects were rather obvious in LOF carriers (−1.00 to 1.00; Fig. 3D). Of note, participants with LOF *ASXL1* or *TET2* variants had a 0.45 or 0.34 decrease in *Z*-adjusted LTL compared with those without any LTL variants, respectively. As the effect of age on LTL was modest (~0.023 SD in decrease per year) [20], the rare-variant genes play an essential role in determining LTL.

Biological function and tissue expression of the rare-variant genes

Bioinformatics analyses of the 18 significant rare-variant genes associated with LTL were conducted to profile their biological functions. Gene Ontology analysis indicated that the genes were mainly enriched in the biological processes associated with telomere maintenance and capping, as well as the molecular functions associated with telomeric and telomerase binding (Fig. 4A). Moreover, the genes were mainly enriched within the cellular component like telomere cap complex (Fig. 4A). Next, we performed protein–protein interaction (PPI) analysis of the significant genes associated with LTL. Within the PPI network, 16 of the 18 genes, including three novel genes (*ASXL1*, *CFAP58*, and *TET2*), were identified, and *TERT*, *TINF2*, and *RTEL1* were identified as the most important genes within the network (Fig. 4B).

Tissue enrichment analysis of the 18 rare-variant genes associated with LTL was performed via GTEx database. There was one gene and two genes enriched in the kidney and testis, respectively (Fig. 4C). *TERT* was specifically enriched in the kidney, while *CFAP58* and *TYMS* were specifically enriched in the testis (Fig. S7). Then, we used scRNA-seq data of human testis to evaluate the expression levels of *CFAP58* and *TYMS* in different cell types within the testis (Fig. 4D). Though the expression level of *CFAP58* was relatively low, it was mainly expressed


max_MAF
■ 0.001
▲ 0.01
◆ 1e−04
● 1e−05
• 5e−06
A
Pan-ancestry
LOF
−log10(Pvalue)
0
20
40
60
POT1
CTC1
RTEL1
TERF1
SAMHD1
ASXL1
ZSWIM3
DCLRE1B
TET2
TERT
HBB
STN1
ATM
PARN
TYMS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 19 21
B
Pan-ancestry
LOF;Missense
−log10(Pvalue)
0
20
40
60
POT1
RTEL1
TERF1
CTC1
SAMHD1
DCLRE1B
TERT
ATM
TET2
ZNF451
CFAP58
HBB
TINF2
PARN
TK1
TYMS
ZSWIM3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 19 21
C
European
LOF
−log10(Pvalue)
0
10
20
30
40
50
POT1
CTC1
TERF1
RTEL1
DCLRE1B
TET2
TERT
STN1
ATM
PARN
TYMS
ASXL1
SAMHD1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 19 21
D
European
LOF;Missense
−log10(Pvalue)
0
10
20
30
40
50
POT1
RTEL1
CTC1
SAMHD1
TERT
DCLRE1B
TERF1
CFAP58
ATM
TET2
ZNF451
TINF2
PARN
TK1
TYMS
ZSWIM3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 19 21

◀**Fig. 2** Exome-wide association analysis of LTL. **A** Manhattan plot showing the results of the rare variants (LOF) from ExWAS of LTL with four different MAF thresholds in pan-ancestry gene-based analysis. **B** Manhattan plot showing the results of the rare variants (LOF and/or missense) from ExWAS of LTL with four different MAF thresholds in pan-ancestry gene-based analysis. **C** Manhattan plot showing the results of the rare variants (LOF) from ExWAS of LTL with four different MAF thresholds in European-specific gene-based analysis. **D** Manhattan plot showing the results of the rare variants (LOF and/or missense) from ExWAS of LTL with four different MAF thresholds in European-specific gene-based analysis. The *x*-axis indicates the position of the gene in 22 chromosomes. The *y*-axis indicates the $-\log10$ of the *p*-value for each association. The black dashed horizontal line indicates the threshold for significant association (p-value $< 2.49 \times 10^{-7}$). The *p*-values are adjusted for age at recruitment, gender, ethnicity, and top 10 principal components. The novel genes identified from ExWAS were marked with red

by spermatids (Fig. 4E). *TYMS* was also expressed by germ cells, and spermatocytes showed the highest level of *TYMS* (Fig. 4E).

Phenotypic associations of the novel rare-variant genes

To explore the underlying pleiotropy of the novel genes significantly associated with LTL, we performed phenotypic association analyses across 17,361 binary (mainly diseases) and 1419 continuous phenotypes from the UK Biobank by AstraZeneca PheWAS [31]. Regarding the phenotypes associated with diseases, our results were consistent with the previous studies, with most of the associated phenotypes deriving from cancer and hematological diseases (Fig. 5A). Of note, *ASXL1* and *TET2* showed significant associations with cancer and hematological diseases, and we observed the largest number of hematological, endocrine and metabolic, and respiratory disease associations for *TET2*. Regarding the associations of the rare-variant genes with continuous biomedical phenotypes (Fig. 5B), *ASXL1* was associated with the largest number of biomedical traits. In line with the obvious associations with hematological and circulatory diseases, the most associations with blood assays and cardiovascular traits were identified for *ASXL1* (Fig. 5B). In addition, *ASXL1* was also associated with most physical measures and pulmonary functions. *TET2* was mainly associated with blood assays (Fig. 5B), supporting its essential role in hematological and immune diseases [47].

In analysis of binary traits, the most significant associations with *ASXL1* were myeloid leukemia and acute myeloid leukemia (Fig. 5C). In addition, significant associations between *ASXL1* and benign or malignant hematological disorders were observed, including myelodysplastic syndromes, diseases of the blood, and other anemias (Fig. 5C). Moreover, *ASXL1* was also significantly associated with hypertension and hypertensive diseases. The most significant associations with *TET2* were cancers of hematopoietic system, including myeloid leukemia, lymphomas, and monocytic leukemia (Fig. 5D). In addition, *TET2* was associated with many hemorrhagic diseases, like purpura and thrombocytopenia. Intriguingly, both *ASXL1* and *TET2* were significantly associated with influenza and pneumonia and renal failure (Fig. 5C, D). No significant associations were observed for *CFAP58* (Table S2).

After multiple comparisons, *TET2* was associated with the largest number of continuous traits, mainly blood assays (Fig. 5E). We found that both *ASXL1* and *TET2* were significantly associated with inflammatory parameters, like white blood cell count, neutrophil count, and monocyte count (Fig. 5E). And no significant associations were observed for *CFAP58*, either (Table S3).

Overall, the analysis of the WES data revealed five novel genes associated with LTL and the biomedical pleiotropies of the genes, with most associations with cancers, hematological diseases, and the related phenotypes particularly blood assays.

Longitudinal disease risk of the novel rare-variant genes

To test the significance of the novel genes in the real-world settings, we further performed survival analyses by Cox proportional hazards models to explore whether the individuals with LTL genes were at a higher risk of incident cancers and other age-related diseases (Table S4). A total of 26 nominal longitudinal associations ($p < 0.05$) were observed for the three genes, where *TET2* showed the largest associations ($n = 19$; Fig. 6A). The complete results of the survival analyses of diseases are shown in Table S5.

In consistent with the results of phenome-wide associations, we found that individuals with *ASXL1* variants had elevated risk of developing leukemia or lymphoma (hazard ratio [HR] = 1.50, 95%

Table 1 Rare-variant gene associations with leukocyte telomere length after Bonferroni corrections

Gene	Group	Pan-ancestry				European			
		p-value	BETA for burden	SE for burden	*p*-value for burden	*p*-value	BETA for burden	SE for burden	*p*-value for burden
DCLRE1B	lof	1.93E−15	0.023	0.003	4.00E−15	3.10E−14	0.025	0.003	1.30E−13
DCLRE1B	missense;lof	3.05E−35	0.022	0.002	2.10E−35	1.05E−27	0.022	0.002	7.33E−28
TET2	lof	5.19E−15	−0.015	0.002	7.41E−16	5.21E−13	−0.015	0.002	5.21E−13
TET2	missense;lof	6.74E−14	−0.014	0.002	9.62E−15	5.83E−12	−0.014	0.002	5.83E−12
TERT	lof	6.69E−13	−0.040	0.006	6.69E−13	2.68E−11	−0.040	0.006	2.68E−11
TERT	missense;lof	5.89E−35	−0.034	0.003	5.89E−35	1.43E−32	−0.035	0.003	1.43E−32
ZNF451	missense;lof	5.82E−10	0.008	0.001	6.69E−10	2.15E−10	0.008	0.001	1.46E−10
POT1	lof	1.74E−65	0.039	0.002	8.32E−60	2.39E−51	0.038	0.003	5.44E−47
POT1	missense;lof	4.83E−44	0.024	0.002	6.90E−45	1.30E−45	0.016	0.001	1.24E−40
TERF1	lof	5.54E−29	0.027	0.003	1.31E−23	3.51E−34	0.042	0.004	5.08E−32
TERF1	missense;lof	7.75E−51	0.027	0.002	3.57E−45	2.73E−54	0.033	0.002	1.17E−47
*CFAP58**	missense;lof	4.88E−14	0.003	0.000	3.34E−07	8.04E−13	0.003	0.001	5.96E−08
STN1	lof	2.07E−09	0.015	0.003	2.89E−09	4.31E−09	0.016	0.003	4.72E−08
ATM	lof	3.89E−10	−0.007	0.001	4.43E−10	1.37E−07	−0.007	0.001	6.43E−08
ATM	missense;lof	5.44E−14	−0.005	0.001	8.13E−15	1.59E−13	−0.004	0.001	3.71E−13
HBB	lof	9.80E−09	0.015	0.003	4.60E−08	1.03E−04	0.012	0.005	1.94E−02
HBB	missense;lof	9.80E−09	0.015	0.003	4.60E−08	1.03E−04	0.012	0.005	1.94E−02
*TINF2***	missense;lof	2.90E−08	0.011	0.003	2.70E−05	8.46E−08	0.022	0.004	8.46E−08
PARN	lof	1.08E−13	−0.027	0.003	1.54E−14	6.85E−13	−0.028	0.004	9.79E−14
PARN	missense;lof	3.37E−08	−0.007	0.001	4.90E−08	9.56E−09	−0.007	0.001	4.65E−09
CTC1	lof	6.95E−17	0.014	0.002	3.29E−17	9.39E−41	0.016	0.001	4.22E−37
CTC1	missense;lof	6.91E−18	0.009	0.001	9.49E−17	1.37E−42	0.011	0.001	5.43E−38
TK1	missense;lof	1.98E−09	−0.012	0.002	2.17E−09	1.97E−07	−0.011	0.002	1.62E−07
TYMS	lof	5.85E−08	−0.035	0.006	5.85E−08	9.18E−08	−0.037	0.007	9.18E−08
TYMS	missense;lof	6.06E−10	−0.016	0.003	1.05E−09	8.15E−09	−0.016	0.003	5.70E−08
ASXL1	lof	5.58E−11	−0.017	0.003	7.97E−12	2.16E−10	−0.017	0.003	1.23E−10
RTEL1	lof	1.12E−35	−0.022	0.002	6.85E−30	1.03E−32	−0.023	0.002	2.33E−26
RTEL1	missense;lof	8.50E−62	−0.016	0.001	9.37E−61	1.23E−47	−0.016	0.001	5.28E−46
SAMHD1	lof	3.15E−12	0.018	0.002	4.50E−13	2.16E−10	0.017	0.003	1.92E−10
SAMHD1	missense;lof	2.00E−41	0.018	0.001	2.85E−42	7.51E−43	0.018	0.001	2.47E−43
*ZSWIM3****	lof	3.68E−08	0.029	0.005	5.03E−08	9.87E−05	0.029	0.007	9.87E−05
ZSWIM3	missense;lof	1.16E−10	0.010	0.002	9.19E−08	2.56E−08	0.024	0.004	2.56E−08

*Max minor allele frequency = 0.01

**Max minor allele frequency = 0.00001

***Max minor allele frequency = 0.000005

confidence interval [CI] 1.12–2.01, $p = 7.86 \times 10^{-3}$) and renal failure (HR = 1.20, 95% CI 1.06–1.36, $p = 4.95 \times 10^{-3}$). *ASXL1* was also associated with the risk of heart failure (HR = 1.28, 95% CI 1.06–1.54, $p = 8.88 \times 10^{-3}$). Regarding *TET2*, we observed significant increased risks of leukemia or lymphoma (HR = 6.95, 95% CI 5.03–9.61, $p = 9.08 \times 10^{-32}$) and any cancers (HR = 1.55, 95% CI 1.27–1.88, $p = 1.22 \times 10^{-5}$). Moreover, *TET2* was also associated with diseases of the blood (HR = 2.35, 95% CI 1.91–2.88, $p = 2.24 \times 10^{-16}$), anemia (HR = 1.61, 95% CI 1.25–2.09, $p = 3.07 \times 10^{-4}$), diseases of the genitourinary system (HR = 1.33, 95% CI 1.09–1.62,

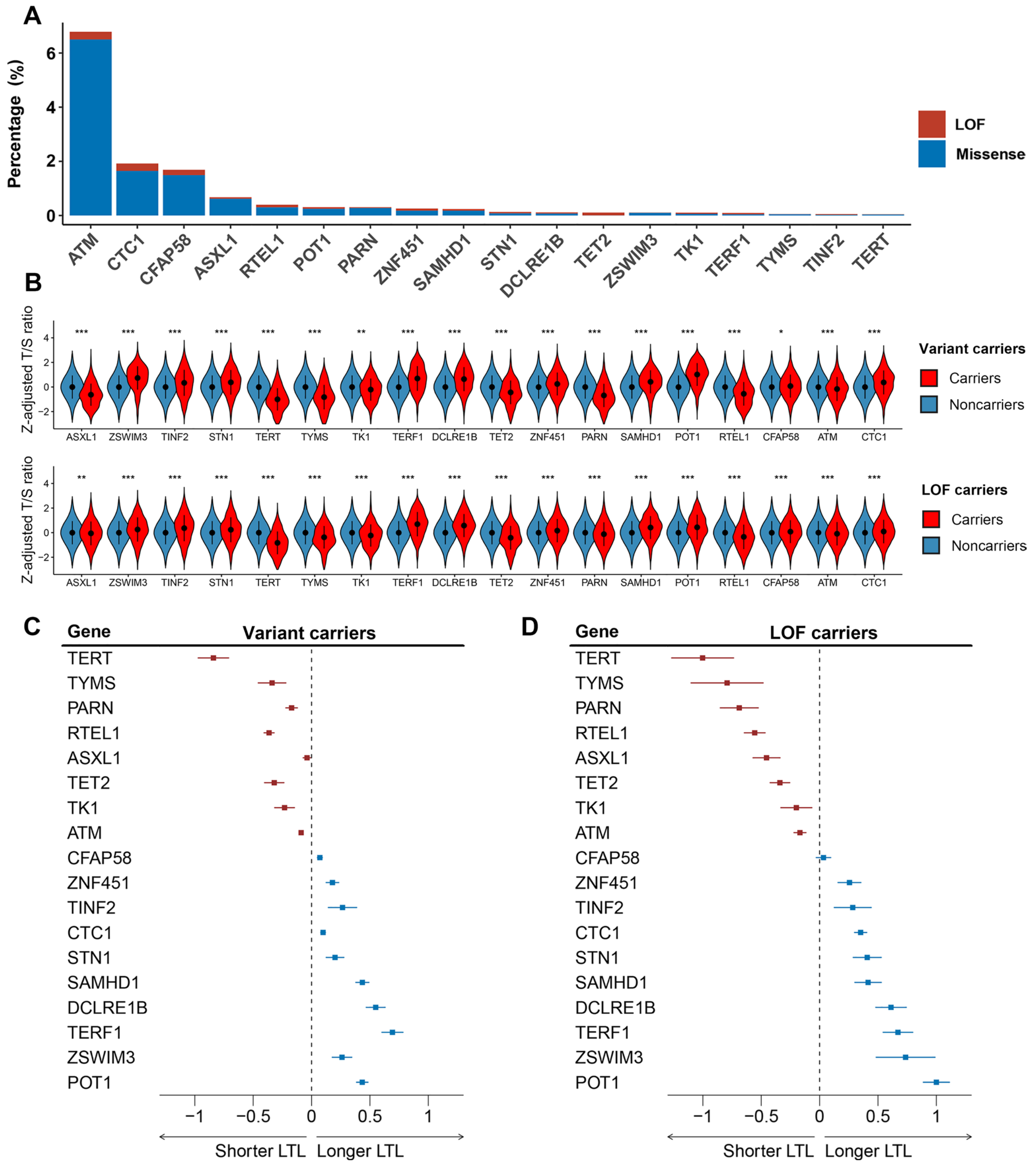


Fig. 3 Landscape of the LTL genes and their effects on LTL. **A** The carrier percentage for rare LOF and missense variants in genes associated with LTL. The percentage was calculated based on all participants in the UK Biobank. The color of the bar indicates LOF (red) or missense (blue) groups of the variants. **B** The LTL measurements are displayed among participants with or without LTL rare gene variants. The violin plot showed the distribution of LTL among the carriers or the non-carriers. The dot represented the median value of LTL, and the line represented 25th and 75th percentiles of LTL. **C** The effects of the genes on LTL among rare gene carriers. The genes are shown on the *y*-axis and the β for the effects on LTL are shown on the *x*-axis. Error bars indicate 95% Cis. **D** The effects of the genes on LTL among LOF rare gene carriers. $*p < 0.05$, $**p < 0.01$; $***p < 0.001$

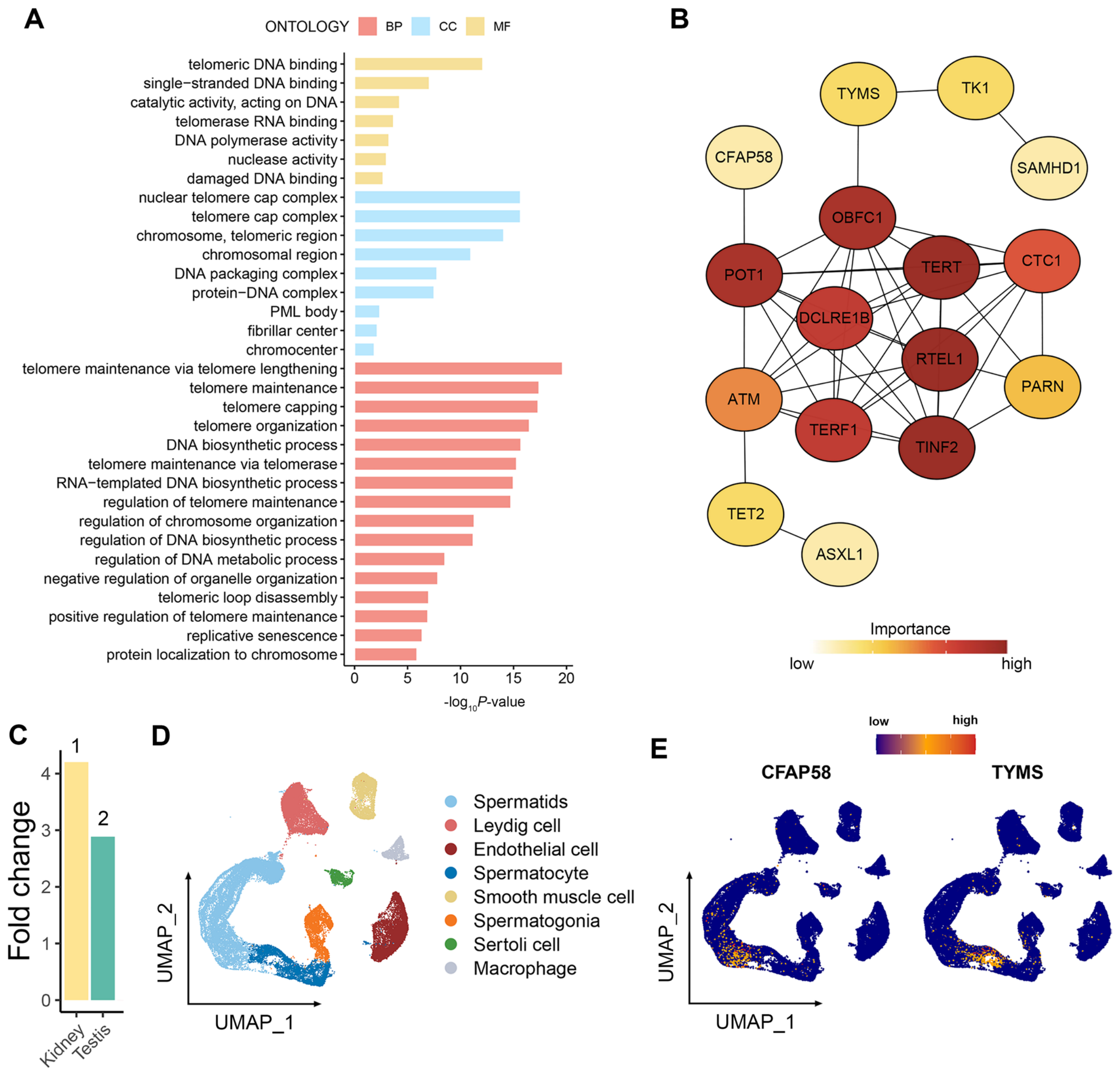


Fig. 4 Bioinformatics analyses of the rare-variant genes associated with LTL. **A** Representative results of the GO analysis. The *x*-axis indicates − log10 of the *p*-value for each GO term. The *y*-axis indicates different GO terms, and each ontology is marked with different colors. **B** Results of the PPI analysis. The color of the gene indicates the importance of the gene within the network. **C** The bar plot showing the tissue-specific gene enrichment. The *x*-axis indicates the types of the tissues. The *y*-axis indicates the fold-change values of the tissue-specific gene enrichment. The number above the bar indicates the number of the tissue-specific genes within the tissue. **D** The uniform manifold approximation and projection (UMAP) plot showing the eight different cell types within the testis. The color of the dots indicates the cell type. **E** The feature plot showing the expression level of *CFAP58* or *TYMS* in different cell types within testis. BP, biological process; CC, cellular component; GO, Gene Ontology; MF, molecular function; PPI, protein–protein interaction

$p = 4.65 \times 10^{-3}$), and renal failure (HR = 1.64, 95% CI 1.30–2.10, $p = 6.27 \times 10^{-5}$), supporting the findings in phenome-wide association analysis. In addition, we also found that *TET2* carriers were associated with diseases of the respirator system (HR = 1.65, 95% CI 1.35–2.10, $p = 8.55 \times 10^{-7}$) and digestive system (HR = 1.36, 95% CI 1.14–1.62, $p = 5.46 \times 10^{-4}$).

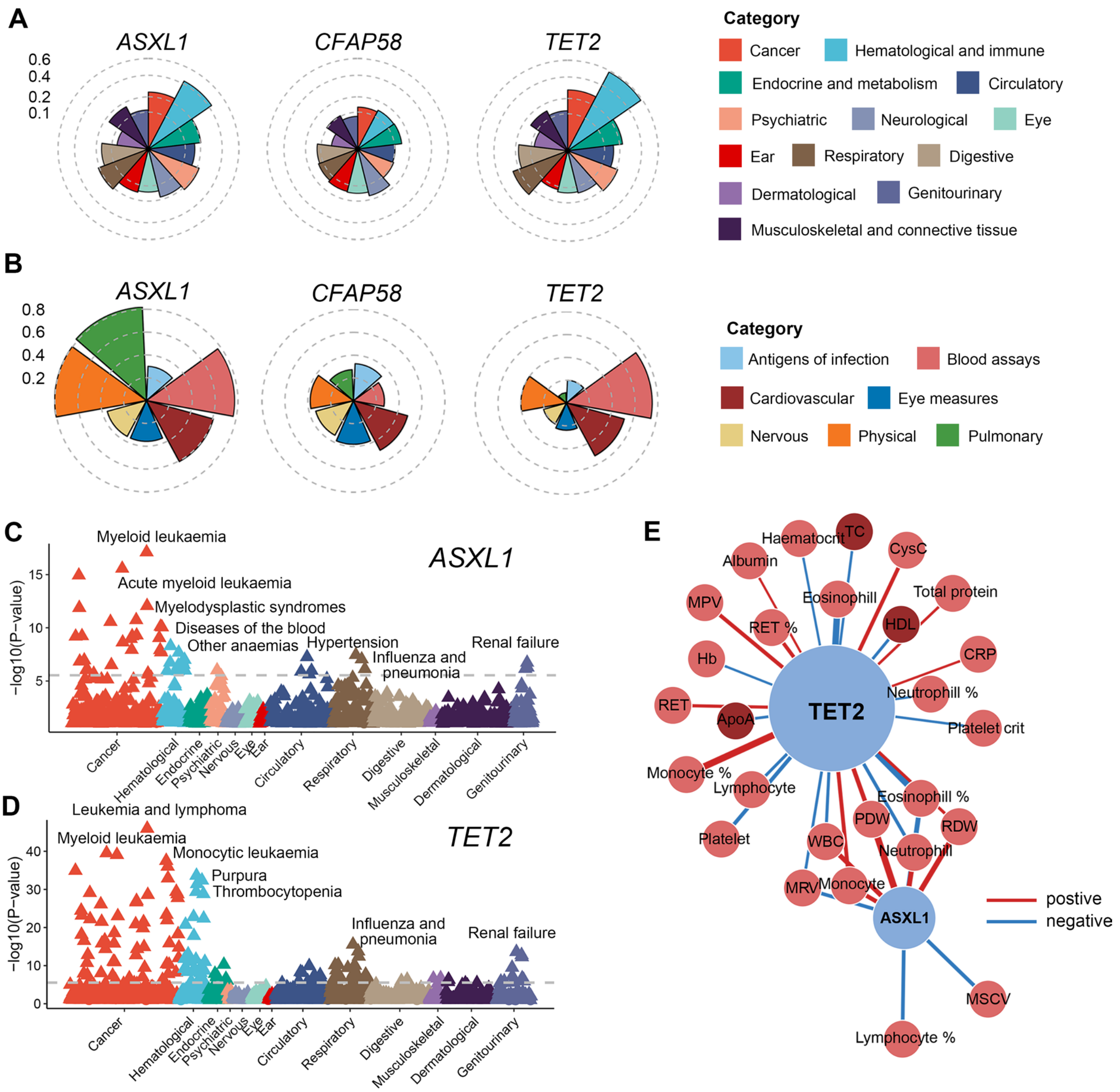
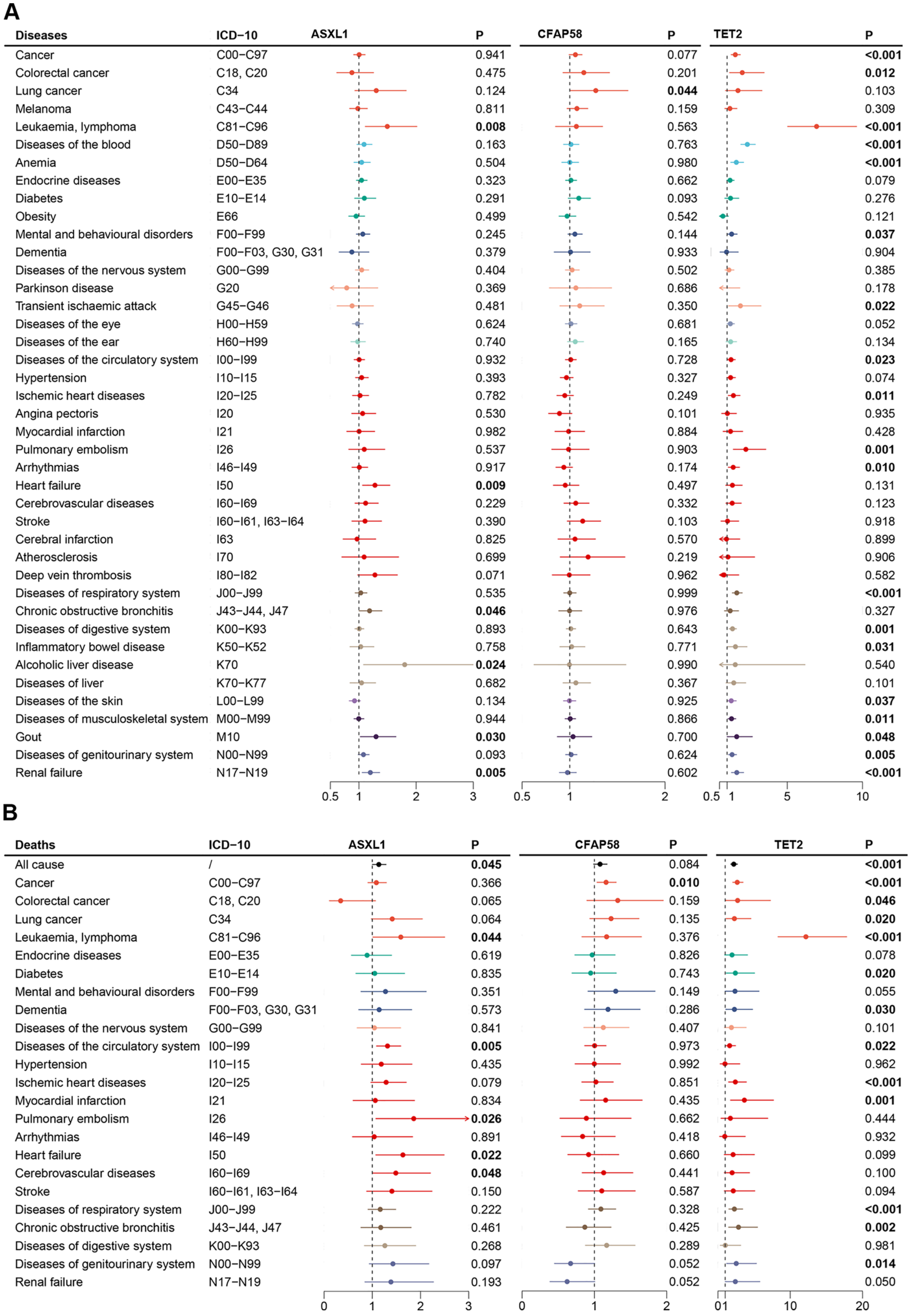


Fig. 5 Phenotypic associations of the rare-variant genes associated with LTL. **A** The associations with diseases. The *y*-axis (concentric circles) indicates the proportion of the phenotypes within a category that are nominally associated (p-value < 0.05) within each gene. **B** The associations with continuous phenotypes. **C** Significant associations of *ASXL1* with binary outcomes. The *x*-axis indicates the binary outcomes. The *y*-axis indicated the − log10 of the *p*-value. The color of the point indicates different categories. The gray dashed horizontal line indicates the threshold for significant association (p-value < 2.88×10^{-6}). **D** Significant associations of *TET2* with binary outcomes. **E** Significant associations of *ASXL1* and *TET2* with continuous traits. The color of the traits indicates different categories. The color of the edge indicates the direction of the association and the width of the edge indicates the − log10 of the *p*-value for each association. ApoA, apolipoprotein A; CRP, C-reactive protein; CysC, cystatin C; Hb, hemoglobin; HDL, high-density lipoprotein; MPV, mean platelet volume; MRV, mean reticulocyte volume; MSCV, mean sphered cell volume; PDW, platelet distribution width; RDW, red blood cell distribution width; RET, reticulocyte count; TC, total cholesterol; WBC, white blood cell count



Diseases	ICD−10	ASXL1	CFAP58	TET2
		P	P	P
Cancer	C00−C97	0.941	0.077	**<0.001**
Colorectal cancer	C18, C20	0.475	0.201	**0.012**
Lung cancer	C34	0.124	**0.044**	0.103
Melanoma	C43−C44	0.811	0.159	0.309
Leukaemia, lymphoma	C81−C96	**0.008**	0.563	**<0.001**
Diseases of the blood	D50−D89	0.163	0.763	**<0.001**
Anemia	D50−D64	0.504	0.980	**<0.001**
Endocrine diseases	E00−E35	0.323	0.662	0.079
Diabetes	E10−E14	0.291	0.093	0.276
Obesity	E66	0.499	0.542	0.121
Mental and behavioural disorders	F00−F99	0.245	0.144	**0.037**
Dementia	F00−F03, G30, G31	0.379	0.933	0.904
Diseases of the nervous system	G00−G99	0.404	0.502	0.385
Parkinson disease	G20	0.369	0.686	0.178
Transient ischaemic attack	G45−G46	0.481	0.350	**0.022**
Diseases of the eye	H00−H59	0.624	0.681	0.052
Diseases of the ear	H60−H99	0.740	0.165	0.134
Diseases of the circulatory system	I00−I99	0.932	0.728	**0.023**
Hypertension	I10−I15	0.393	0.327	0.074
Ischemic heart diseases	I20−I25	0.782	0.249	**0.011**
Angina pectoris	I20	0.530	0.101	0.935
Myocardial infarction	I21	0.982	0.884	0.428
Pulmonary embolism	I26	0.537	0.903	**0.001**
Arrhythmias	I46−I49	0.917	0.174	**0.010**
Heart failure	I50	**0.009**	0.497	0.131
Cerebrovascular diseases	I60−I69	0.229	0.332	0.123
Stroke	I60−I61, I63−I64	0.390	0.103	0.918
Cerebral infarction	I63	0.825	0.570	0.899
Atherosclerosis	I70	0.699	0.219	0.906
Deep vein thrombosis	I80−I82	0.071	0.962	0.582
Diseases of respiratory system	J00−J99	0.535	0.999	**<0.001**
Chronic obstructive bronchitis	J43−J44, J47	**0.046**	0.976	0.327
Diseases of digestive system	K00−K93	0.893	0.643	**0.001**
Inflammatory bowel disease	K50−K52	0.758	0.771	**0.031**
Alcoholic liver disease	K70	**0.024**	0.990	0.540
Diseases of liver	K70−K77	0.682	0.367	0.101
Diseases of the skin	L00−L99	0.134	0.925	**0.037**
Diseases of musculoskeletal system	M00−M99	0.944	0.866	**0.011**
Gout	M10	**0.030**	0.700	**0.048**
Diseases of genitourinary system	N00−N99	0.093	0.624	**0.005**
Renal failure	N17−N19	**0.005**	0.602	**<0.001**

Deaths	ICD−10	ASXL1	CFAP58	TET2
		P	P	P
All cause	/	**0.045**	0.084	**<0.001**
Cancer	C00−C97	0.366	**0.010**	**<0.001**
Colorectal cancer	C18, C20	0.065	0.159	**0.046**
Lung cancer	C34	0.064	0.135	**0.020**
Leukaemia, lymphoma	C81−C96	**0.044**	0.376	**<0.001**
Endocrine diseases	E00−E35	0.619	0.826	0.078
Diabetes	E10−E14	0.835	0.743	**0.020**
Mental and behavioural disorders	F00−F99	0.351	0.149	0.055
Dementia	F00−F03, G30, G31	0.573	0.286	**0.030**
Diseases of the nervous system	G00−G99	0.841	0.407	0.101
Diseases of the circulatory system	I00−I99	**0.005**	0.973	**0.022**
Hypertension	I10−I15	0.435	0.992	0.962
Ischemic heart diseases	I20−I25	0.079	0.851	**<0.001**
Myocardial infarction	I21	0.834	0.435	**0.001**
Pulmonary embolism	I26	**0.026**	0.662	0.444
Arrhythmias	I46−I49	0.891	0.418	0.932
Heart failure	I50	**0.022**	0.660	0.099
Cerebrovascular diseases	I60−I69	**0.048**	0.441	0.100
Stroke	I60−I61, I63−I64	0.150	0.587	0.094
Diseases of respiratory system	J00−J99	0.222	0.328	**<0.001**
Chronic obstructive bronchitis	J43−J44, J47	0.461	0.425	**0.002**
Diseases of digestive system	K00−K93	0.268	0.289	0.981
Diseases of genitourinary system	N00−N99	0.097	0.052	**0.014**
Renal failure	N17−N19	0.193	0.052	0.050

◂**Fig. 6** The associations of LTL rare-variant gene carriers with **A** incident diseases and **B** deaths. The forest plot showing the results of the Cox proportional hazard models of the risks of diseases and deaths among the carriers with LTL rare-variant genes. The *y*-axis indicates the diseases. The disease with a *p*-value < 0.05 was marked with bold in the plots

Longitudinal death risk of the novel rare-variant genes

We also performed survival analyses to explore whether the individuals carrying LTL rare-variant genes were at a higher risk of overall deaths or cause-specific deaths (Fig. 6B). The complete results of the survival analyses of deaths are shown in Table S6.

We found that *TET2* carriers had an increased risk of all-cause death (HR = 2.19, 95% CI 1.78–2.70, $p = 1.57 \times 10^{-13}$), death due to any cancers (HR = 2.67, 95% CI 2.05–3.47, $p = 2.34 \times 10^{-13}$), and leukemia or lymphoma (HR = 12.17, 95% CI 8.31–17.80, $p = 7.35 \times 10^{-38}$). Moreover, there is significant longitudinal risk of *TET2* with ischemic heart disease (HR = 2.41, 95% CI 1.47–3.93, $p = 4.55 \times 10^{-4}$) and myocardial infarction (HR = 3.66, 95% CI 1.74–7.69, $p = 6.21 \times 10^{-4}$). In line with the longitudinal disease risk, *TET2* carriers had a higher risk of diseases of the respiratory system (HR = 2.27, 95% CI 1.51–3.42, $p = 8.52 \times 10^{-5}$). *ASXL1* carriers probably had an increased risk of all-cause death (HR = 1.14, 95% CI 1.00–1.29, $p = 4.45 \times 10^{-2}$) and death due to circulatory diseases (HR = 1.32, 95% CI 1.09–1.59, $p = 4.59 \times 10^{-3}$). However, we failed to observe increased risk of all-cause death among the carriers with *CFAP58* variants ($p > 0.05$).

Overall, through survival analyses, we found that the novel LTL rare-variant genes were associated with multiple adverse outcomes. The participants carrying *TET2* were at higher risk of hematological malignancies and all-cause deaths, and those with *ASXL1* variants were at higher risk of circulatory diseases.

Mendelian randomization of LTL with biomedical phenotypes

Next, to support the identified rare variant associations of LTL and explore the common variant associations, we performed GWAS analysis of LTL in the European population (Fig. S8) and identified 231 independent variants and 101 genomic risk loci for LTL at genome-wide significance ($p < 5 \times 10^{-8}$; Table S7). The estimated effects of the significant variants were generally modest (most absolute effects were less than 0.1 SD per allele), which was consistent with the previous GWAS of LTL (Fig. S8B and S8C) [10].

Then, we performed MR analyses to evaluate the causal associations of LTL with multiple biomedical phenotypes at genetic level. A total of 186 traits were screened (Fig. S9; Table S8). The results of the forward (LTL as the exposure) MR analyses were consistent with the phenotypic associations identified in the ExWAS analyses (Fig. S10A); our analysis indicated that genetically predicted LTL was positively associated with hematological cancer ($\beta = 0.003$, standard error (SE) = 0.001, $p = 0.046$). In addition to cancers, genetically predicted LTL was negatively associated with several aging-related diseases, like CAD ($\beta = -0.131$, SE = 0.043, $p = 0.002$), idiopathic pulmonary fibrosis ($\beta = -0.003$, SE = 0.001, $p = 2.10 \times 10^{-5}$), and family history of AD ($\beta = -0.012$, SE = 0.004, $p = 0.001$). Regarding the backward MR analysis of LTL (Fig. S10B), genetically predicted parental longevity was associated with LTL (combined parental attained age: $\beta = 0.155$, standard error (SE) = 0.071, $p = 0.029$; mother's attained age: $\beta = 0.231$, SE = 0.055, $p = 2.33 \times 10^{-5}$). Moreover, we also identified several peripheral biomarkers associated with LTL, including low-density lipoprotein cholesterol (LDL-C, $\beta = 0.017$, SE = 0.006, $p = 0.004$) and white blood cell count (WBC, $\beta = -0.020$, SE = 0.010, $p = 0.050$). The full MR analyses have been presented in Table S9. Overall, the MR analyses further supported the associations between LTL and hematological malignancies and cardiovascular diseases and revealed some traits associated with LTL at genetic level.

Discussion

To our best knowledge, the present study was the largest ExWAS study of LTL, which systematically elucidated the rare genetic determinants of LTL and their biomedical implications. In ExWAS analyses of LTL, we replicated many known telomere syndrome–causing variants and discovered three novel genes not previously related to telomere length or telomere syndrome that serve as promising candidates for further experimental investigations. We characterized their biological functions associated with telomere maintenance and capping and the specific tissue expression in the testis and profiled considerable phenotypic associations of *ASXL1* and *TET2* with diseases including cancers and hematological, cardiovascular, and genitourinary diseases, and physiological traits including blood assays and cardiovascular traits. Survival analyses further supported their associations with hematological malignancies and age-related disorders. Overall, our ExWAS analyses of LTL revealed novel genes associated with LTL which served as promising candidates for further experimental investigations, and provided insights into the associations of LTL with clonal hematopoiesis and age-related disorders at genetics level.

Genetics is essential in telomere length and telomere syndrome [48], which explained approximately 70% of the variance in LTL [49]. However, the effect sizes of the genes or other exposures on LTL reported in previous studies were generally modest. Codd et al. demonstrated that the estimated effect sizes of the common variants on LTL were moderate (<0.2 SD per allele) [10]. Meanwhile, though LTL could be modified by healthy behaviors to some extent, their effects on LTL were moderate, accounting for less than 0.2% of the variation [20]. Our findings revealed their significant effect on LTL compared with the common variants. For instance, participants with *TERT* LOF variants had a 1.00 decrease in *Z*-adjusted log LTL. And the estimated effects for *ASXL1* and *TET2* LOF variants were − 0.45 and − 0.34 on *Z*-adjusted log LTL, respectively, equivalent to ~ 15–20 years of age-related change in LTL [20]. These highlighted the putative role of rare-variant genes in LTL and the related consequences and helped the identification of those aged and at higher risk of age-related disorders.

Previous studies have demonstrated that mutations in telomere and telomerase genes contributed to telomere shortening that manifests in age-related phenotypes, providing evidence that telomeres were associated with aging [50, 51]. Our study has replicated several known genes implicated in monogenic telomere disorders [51]. *POT1* was the most significant gene associated with LTL in our ExWAS analysis, which encoded the subunit of telomeric structure [52]. Previous studies have shown the protective role of POT1 protein in DNA damage response [52] and that *POT1* mutations led to telomere dysfunction [53]. Compared with previous rare and ultra-rare variant ExWAS study based on 200 k exome sequencing data [10], our study significantly expanded the sample size, using the latest 450 k exome sequencing data in the UK Biobank, replicated several LTL genes (i.e., *RTEL1*, *TERF1*, and *TERT*), and further identified three novel genes associated with LTL. Compared with a recent ExWAS study of LTL [46], our study has used SAIGE-GENE + gene–based analysis, which improved the statistical power and robustness of results, identified novel gene-LTL associations, and further characterized the biomedical implications of the novel rare-variant genes. *TET2* encodes an essential enzyme that catalyzes oxidative responses of methylcytosine bases [54]. Among individuals older than 65 years old, 5–10% of them showed inactivating mutations within *TET2* in peripheral blood cells [55]. The biological functions of *TET2* involved with the self-renewal and differentiation of hematopoietic stem cell, maintenance of genomic stability, and participating inflammatory responses [56]. Yang et al. demonstrated that knockout of TET enzymes (*TET1/2/3*) in embryonic stem cells induced shorter LTL and instability of chromosome by modulating the expression of *DNMT3A* and methylation [57]. Mutations of *TET2* contributed to higher levels of inflammatory molecules, which further recruit circulating immune cells and form a chronic inflammatory microenvironment, thus accelerating cellular senescence or leading to age-related diseases [58]. In addition to the known association of *TET2* with hematological malignancies [59], we revealed putative associations between *TET2* and diseases of hematological, circulatory, respiratory, and genitourinary systems. Such associations were supported by longitudinal risk among *TET2* carriers and the phenotypic associations of *TET2* with related biomedical traits. For instance, *TET2* carriers had a higher risk of developing diseases involving blood and any anemias and *TET2* was significantly associated with lower hemoglobin and platelet count. In addition, *ASXL1* is an epigenetic modulator that frequently mutates in myeloid neoplasms, such as myeloid leukemia, myelodysplastic syndromes, and myeloproliferative neoplasms [60]. Mutant *ASXL1* induced dysfunction and age-related aberrant proliferation of hematopoietic stem cells and lymphocytes, thus promoting cellular senescence and the progression of malignant tumors

[61, 62]. A recent retrospective study suggested that among subjects with acute myeloid leukemia and *ASXL1* mutations, the risk of developing cardiovascular events increased by 40% compared with the noncarriers [62]. Moreover, our study suggested that *ASXL1* carriers had increased risk of heart failure and deaths due to circulatory diseases, consistent with a recent cohort study by Yu et al., demonstrating that CHIP due to *ASXL1* mutations was associated with increasing risk of incident heart failure and decreasing left ventricular ejection fraction [63]. Similar to *TET2*, *ASXL1*-mediated clonal hematopoiesis induced a pro-inflammatory phenotype of cardiac macrophages characterized by higher expression levels of interleukin-1β (IL-1β) and IL-6, further forming a pro-inflammatory environment in heart and accelerating heart failure [64]. Another key finding of our analysis was that both *ASXL1* and *TET2* were associated with age-related disorders of multiple systems, including circulatory, respiratory, and genitourinary systems. The associations could be explained by their impact on immune functions. In murine models, knockout of *Tet2* showed increased expression of inflammatory cytokines contributing to atherosclerosis and pulmonary emphysema [65, 66]. Therefore, our results added the phenotypic associations of LTL with age-related disorders at genetic level and implied the potential clinical significance of early screening and management for cancers and other age-related disorders in individuals with *TET2* and *ASXL1* mutations.

Intriguingly, both *ASXL1* and *TET2* were epigenetic modulators that were common among subjects with clonal hematopoiesis of indeterminate potential (CHIP) [46, 67]. And both telomere shortening and CHIP are commonly observed in older individuals and considered hallmarks of aging [68, 69]. Shorter TL was observed in CHIP carriers [70, 71]. GWAS analyses have reported that *TERT* was the common locus shared by CHIP and LTL [10, 72]. By analyzing CHIP with VAF > 10% in Trans-Omics for Precision Medicine (TOPMed) and UK Biobank, Nakao et al. demonstrated the bi-directional associations between CHIP and shorter LTL [73]. In addition to *TERT*, our ExWAS analyses revealed novel LTL genes, *ASXL1* and *TET2*, which may explain the complex relationship between CHIP and telomere length. As the variant allele frequency (VAF) of *ASXL1* and *TET2* was skewed away from 0.5 (~ 0.25) in previous studies utilizing exome sequencing data of the UK Biobank [72, 74], our results supported that clonal hematopoiesis is a putative contributor to the genetic underpinnings of LTL. In addition, we conducted survival analyses of *ASXL1* and *TET2*, which filled the gap regarding the associations among LTL, CHIP, and cancers or age-related disorders.

Regarding one novel LTL gene showing putative protective effects on telomere shortening, *CFAP58* was a protein expressed predominantly in sperms and ciliated cells [75]. *CFAP58* was essential for the ciliogenesis and flagellar elongation, and its mutations have been observed in patients with morphological abnormalities of the sperm flagella or sperm motility disorders [76, 77]. Similarly, tissue enrichment analysis of the rare-variant genes for LTL indicated that some genes associated with LTL were specifically expressed and enriched within testis. Our results were supported by the previous studies which reported that relative TL was the longest in the testis and the expression of TL maintenance enzyme was also the highest in the testis [69]. *CFAP58* was associated with ferroptosis and immune infiltration in cancers and showed predictive values in the prognosis of endometrial carcinoma [78, 79], which supported the suggestive association between *CFAP58* carriers and deaths due to cancers. However, the underlying cellular and molecular mechanisms of *CFAP58* with LTL or aging need further investigations.

There were several limitations in the present study. Firstly, both *ASXL1* and *TET2* are essential CHIP genes and the somatic mutations in the blood cells may lead to contamination in rare variant identification. And we have checked that the individuals with *ASXL1* or *TET2* variants rarely experienced hematological malignancies (less than 5 in *ASXL1* or *TET2* carriers at baseline). And our analyses have adjusted age at recruitment as a covariate to avoid the impact of age on the results. Secondly, it should be noted that the participants in the UK Biobank are mainly middle-aged, and our analysis was restricted to those of European ancestry. Therefore, genetic determinants of LTL in different age strata and ancestries need to be further investigated. Thirdly, the results of our analysis lack an external validation. However, we have adjusted some essential covariates in the association analyses and used multiple models in gene-based collapsing analysis to support the robustness of the results. We also performed additional bioinformatic analyses to support the biological relevance of the genes identified here. Lastly, experimental and functional studies are warranted to assess the functions and mechanisms of the genes in the associations with the shortening or variations of LTL and to test their values as therapeutic targets for age-related diseases.

In conclusion, the present WES study has profiled the landscape of the rare variants with relevance to LTL and the biomedical phenotypes, clonal hematopoiesis, or health-related outcomes associated with LTL. The understanding of LTL would furnish novel insights into the molecular mechanisms and therapeutic targets for aging and age-related diseases.

Acknowledgements We gratefully thank all the participants and researchers from the UK Biobank and MR-Base. We gratefully thank Nie et al. for providing scRNA-seq matrix.

Author contribution All authors had full access to the data in the study and accept responsibility to submit for publication. J. T. Yu and W. Cheng designed the study. W. S. Liu, B. S. Wu, and L. Yang conducted the main analyses and drafted the manuscript. Y. T. Deng, X. R. Wu, and X. Y. He contributed to data collection and analyses. S. D. Chen, Y. R. Zhang, and J. Yang contributed to data interpretation. J. T. Yu, Y. M. Xu, W. Cheng, and J. F. Feng critically revised the manuscript. All authors reviewed and approved the final version.

Funding This study was supported by grants from the Science and Technology Innovation 2030 Major Projects (2022ZD0211600), National Natural Science Foundation of China (92249305, 82271471, 82071201, 82071997), Shanghai Municipal Science and Technology Major Project (2018SHZDZX01), Research Start-up Fund of Huashan Hospital (2022QD002), Excellence 2025 Talent Cultivation Program at Fudan University (3030277001), Shanghai Talent Development Funding for The Project (2019074), Shanghai Rising-Star Program (21QA1408700), 111 Project (B18015), the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2020-PT310-01), and ZHANGJIANG LAB, Tianqiao and Chrissy Chen Institute, the State Key Laboratory of Neurobiology and Frontiers Center for Brain Science of Ministry of Education, Shanghai Center for Brain Science and Brain-Inspired Technology, Fudan University.

Data availability The main data, including individual-level phenotype and sequencing data, used in this study were accessed from the UK Biobank under the application number 19542. The summary-level GWAS data used in the present study were acquired from MR-Base database (https://gwas.mrcieu.ac.uk/datasets/). The scRNA-seq data used in the present study were acquired from GEO database (https://www.ncbi.nlm.nih.gov/geo/) with the accession number: GSE182786.

Code availability The code used for the ExWAS analyses was an adaptation of the R package SAIGE-GENE+and was available through the GitHub repository: https://github.com/saigegit/SAIGE. Gene Ontology analysis was performed via the R package clusterProfiler (v.4.4.4; https://github.com/YuLab-SMU/clusterProfiler) and tissue enrichment analysis was performed via the R package TissueEnrich (v.1.16.0, https://github.com/Tuteja-Lab/TissueEnrich). The code for the main analysis and visualization of single-cell RNA sequencing data was an adaptation of the R package Seurat (v.4.3.0) and was available through the website: https://satijalab.org/seurat/index.html. The code for Mendelian randomization was an adaptation of the R package TwoSampleMR (v.0.5.6) and was available through the GitHub repository: https://github.com/MRCIEU/TwoSampleMR.

Declarations

Ethics approval UK Biobank received ethical approval from the North West Center for Research Ethics Committee (11/NW/0382). And the present study was approved by UK Biobank under the application number 19542.

Competing interests The authors declare no competing interests.

References

1. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. Nature. 1990;345:458–60.
2. Rossiello F, Jurk D, Passos JF, d'Adda di Fagagna F. Telomere dysfunction in ageing and age-related diseases. Nat Cell Biol. 2022;24:135–47.
3. Chakravarti D, LaBella KA, DePinho RA. Telomeres: history, health, and hallmarks of aging. Cell. 2021;184:306–22.
4. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: an expanding universe. Cell. 2022;S0092–8674(22):01377.
5. Whittemore K, Vera E, Martínez-Nevado E, Sanpera C, Blasco MA. Telomere shortening rate predicts species life span. Proc Natl Acad Sci U S A. 2019;116:15122–7.
6. Brouilette SW, Moore JS, McMahon AD, Thompson JR, Ford I, Shepherd J, et al. Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. Lancet. 2007;369:107–14.
7. Kuo C-L, Pilling LC, Kuchel GA, Ferrucci L, Melzer D. Telomere length and aging-related outcomes in humans: a Mendelian randomization study in 261,000 older participants. Aging Cell. 2019;18:e13017.
8. Stuart BD, Choi J, Zaidi S, Xing C, Holohan B, Chen R, et al. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. Nat Genet. 2015;47:512–7.

9. Broer L, Codd V, Nyholt DR, Deelen J, Mangino M, Willemsen G, 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. 2013;21:1163–8.
10. Codd V, Wang Q, Allara E, Musicha C, Kaptoge S, Stoma S, et al. Polygenic basis and biomedical consequences of telomere length variation. Nat Genet. 2021;53:1425–33.
11. Taub MA, Conomos MP, Keener R, Iyer KR, Weinstock JS, Yanek LR, et al. Genetic determinants of telomere length from 109,122 ancestrally diverse whole-genome sequences in TOPMed. Cell Genom. 2022;2:100084.
12. van der Spek A, Warner SC, Broer L, Nelson CP, Vojinovic D, Ahmad S, et al. Exome sequencing analysis identifies rare variants in ATM and RPL8 that are associated with shorter telomere length. Front Genet. 2020;11:337.
13. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. Nature. 2021;599:628–34.
14. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40:695–701.
15. Akbari P, Gilani A, Sosina O, Kosmicki JA, Khrimian L, Fang Y-Y, et al. Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity. Science. 2021;373:eabf8683.
16. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet. 2005;37:161–5.
17. Stein EA, Mellis S, Yancopoulos GD, Stahl N, Logan D, Smith WB, et al. Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. N Engl J Med. 2012;366:1108–18.
18. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779.
19. Codd V, Denniff M, Swinfield C, Warner SC, Papakonstantinou M, Sheth S, et al. A major population resource of 474,074 participants in UK Biobank to investigate determinants and biomedical consequences of leukocyte telomere length. Preprint at medRxiv. 2021;2021.03.18.21253457.
20. Bountziouka V, Musicha C, Allara E, Kaptoge S, Wang Q, Angelantonio ED, et al. Modifiable traits, healthy behaviours, and leukocyte telomere length: a population-based study in UK Biobank. Lancet Healthy Longev. 2022;3:e321–31.
21. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. Nature. 2020;586:749–56.
22. Zhou W, Bi W, Zhao Z, Dey KK, Jagadeesh KA, Karczewski KJ, et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. Nat Genet. 2022;54:1466–9.
23. Zhou W, Zhao Z, Nielsen JB, Fritsche LG, LeFaive J, Gagliano Taliun SA, et al. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. Nat Genet. 2020;52:634–9.
24. Li X, Li Z, Zhou H, Gaynor SM, Liu Y, Chen H, et al. Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. Nat Genet. 2020;52:969–83.
25. Liu Y, Xie J. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. J Am Stat Assoc. 2020;115:393–402.
26. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6:80–92.
27. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc. 2016;11:1–9.
28. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7.20.
29. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res. 2009;19:1553–61.
30. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010;7:575–6.
31. Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. Nature. 2021;597:527–32.
32. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2:100141.
33. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
34. Jain A, Tuteja G. TissueEnrich: tissue-specific gene enrichment analysis. Bioinformatics. 2019;35:1966–7.
35. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Proteomics, et al. Tissue-based map of the human proteome. Science. 2015;347:1260419.
36. Nie X, Munyoki SK, Sukhwani M, Schmid N, Missel A, Emery BR, et al. Single-cell analysis of human testis aging and correlation with elevated body mass index. Dev Cell. 2022;57:1160-1176.e5.
37. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–20.
38. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96.
39. Zhang Z. Survival analysis in the presence of competing risks. Ann Transl Med. 2017;5:47.
40. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9.

41. Ge Y-J, Wu B-S, Zhang Y, Chen S-D, Zhang Y-R, Kang J-J, et al. Genetic architectures of cerebral ventricles and their overlap with neuropsychiatric traits. Nat Hum Behav. 2024;8:164–80.
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
43. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.
44. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512–25.
45. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40:304–14.
46. Kessler MD, Damask A, O'Keeffe S, Banerjee N, Li D, Watanabe K, et al. Common and rare variant associations with clonal haematopoiesis phenotypes. Nature. 2022;612:301–9.
47. Belizaire R, Wong WJ, Robinette ML, Ebert BL. Clonal haematopoiesis and dysregulation of the immune system. Nat Rev Immunol. 2023;23:595–610.
48. DeBoy EA, Tassia MG, Schratz KE, Yan SM, Cosner ZL, McNally EJ, et al. Familial clonal hematopoiesis in a long telomere syndrome. N Engl J Med. 2023;388:2422–33.
49. Vasa-Nicotera M, Brouilette S, Mangino M, Thompson JR, Braund P, Clemitson J-R, et al. Mapping of a major locus that determines telomere length in humans. Am J Hum Genet. 2005;76:147–51.
50. Armanios M. Telomeres and age-related disease: how telomere biology informs clinical paradigms. J Clin Invest. 2013;123:996–1002.
51. Armanios M, Blackburn EH. The telomere syndromes. Nat Rev Genet. 2012;13(10):693–704.
52. Wang F, Podell ER, Zaug AJ, Yang Y, Baciu P, Cech TR, et al. The POT1-TPP1 telomere complex is a telomerase processivity factor. Nature. 2007;445:506–10.
53. Ramsay AJ, Quesada V, Foronda M, Conde L, Martínez-Trillos A, Villamor N, et al. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. Nat Genet. 2013;45:526–30.
54. Cong B, Zhang Q, Cao X. The function and regulation of TET2 in innate immunity and inflammation. Protein Cell. 2021;12:165–73.
55. Ferrone CK, Blydt-Hansen M, Rauh MJ. Age-associated TET2 mutations: common drivers of myeloid dysfunction, cancer and cardiovascular disease. Int J Mol Sci. 2020;21:626.
56. Kunimoto H, Nakajima H. TET2: a cornerstone in normal and malignant hematopoiesis. Cancer Sci. 2021;112:31–40.
57. Yang J, Guo R, Wang H, Ye X, Zhou Z, Dan J, et al. Tet enzymes regulate telomere maintenance and chromosomal stability of mouse ESCs. Cell Rep. 2016;15:1809–21.
58. Jaiswal S, Libby P. Clonal haematopoiesis: connecting ageing and inflammation in cardiovascular disease. Nat Rev Cardiol. 2020;17:137–44.
59. Cimmino L, Dolgalev I, Wang Y, Yoshimi A, Martin GH, Wang J, et al. Restoration of TET2 function blocks aberrant self-renewal and leukemia progression. Cell. 2017;170:1079-1095.e20.
60. Asada S, Fujino T, Goyama S, Kitamura T. The role of ASXL1 in hematopoiesis and myeloid malignancies. Cell Mol Life Sci. 2019;76:2511–23.
61. Fujino T, Goyama S, Sugiura Y, Inoue D, Asada S, Yamasaki S, et al. Mutant ASXL1 induces age-related expansion of phenotypic hematopoietic stem cells through activation of Akt/mTOR pathway. Nat Commun. 2021;12:1826.
62. Calvillo-Argüelles O, Schoffel A, Capo-Chichi J-M, Abdel-Qadir H, Schuh A, Carrillo-Estrada M, et al. Cardiovascular disease among patients with AML and CHIP-related mutations. JACC CardioOncol. 2022;4:38–49.
63. Yu B, Roberts MB, Raffield LM, Zekavat SM, Nguyen NQH, Biggs ML, et al. Supplemental association of clonal hematopoiesis with incident heart failure. J Am Coll Cardiol. 2021;78:42–52.
64. Min K-D, Polizio AH, Kour A, Thel MC, Walsh K. Experimental ASXL1-mediated clonal hematopoiesis promotes inflammation and accelerates heart failure. J Am Heart Assoc. 2022;11:e026154.
65. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. N Engl J Med. 2017;377:111–21.
66. Miller PG, Qiao D, Rojas-Quintero J, Honigberg MC, Sperling AS, Gibson CJ, et al. Association of clonal hematopoiesis with chronic obstructive pulmonary disease. Blood. 2022;139:357–68.
67. Steensma DP. Clinical consequences of clonal hematopoiesis of indeterminate potential. Blood Adv. 2018;2:3404–10.
68. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371:2488–98.
69. Demanelis K, Jasmine F, Chen LS, Chernoff M, Tong L, Delgado D, et al. Determinants of telomere length across human tissues. Science. 2020;369:eaaz6876.
70. Nachun D, Lu AT, Bick AG, Natarajan P, Weinstock J, Szeto MD, et al. Clonal hematopoiesis associated with epigenetic aging and clinical outcomes. Aging Cell. 2021;20:e13366.
71. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood. 2017;130:742–52.
72. Kar SP, Quiros PM, Gu M, Jiang T, Mitchell J, Langdon R, et al. Genome-wide analyses of 200,453 individuals yield new insights into the causes and consequences of clonal hematopoiesis. Nat Genet. 2022;54:1155–66.
73. Nakao T, Bick AG, Taub MA, Zekavat SM, Uddin MM, Niroula A, et al. Mendelian randomization supports bidirectional causality between telomere length and clonal

hematopoiesis of indeterminate potential. Sci Adv. 2022;8:eabl6579.

74. Liu JZ, Chen C-Y, Tsai EA, Whelan CD, Sexton D, John S, et al. The burden of rare protein-truncating genetic variants on human lifespan. Nat Aging. 2022;2:289–94.
75. Li Z-Z, Zhao W-L, Wang G-S, Gu N-H, Sun F. The novel testicular enrichment protein Cfap58 is required for Notch-associated ciliogenesis. Biosci Rep. 2020;40:BSR20192666.
76. Sha Y, Sha Y, Liu W, Zhu X, Weng M, Zhang X, et al. Biallelic mutations of CFAP58 are associated with multiple morphological abnormalities of the sperm flagella. Clin Genet. 2021;99:443–8.
77. Oud MS, Houston BJ, Volozonoka L, Mastrorosa FK, Holt GS, Alobaidi BKS, et al. Exome sequencing reveals variants in known and novel candidate genes for severe sperm motility disorders. Hum Reprod. 2021;36:2597–611.
78. Gu C, Lin C, Zhu Z, Hu L, Wang F, Wang X, et al. The IFN-γ-related long non-coding RNA signature predicts prognosis and indicates immune microenvironment infiltration in uterine corpus endometrial carcinoma. Front Oncol. 2022;12:955979.
79. Qin A, Qian Q, Cui X, Bai W. Ferroptosis-related lncRNA model based on CFAP58-DT for predicting prognosis and immunocytes infiltration in endometrial cancer. Ann Transl Med. 2023;11:151.

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