



Integrated genetic analyses revealed novel human longevity loci and reduced risks of multiple diseases in a cohort study of 15,651 Chinese individuals

Xiaomin Liu^{1,2,3} | Zijun Song⁴ | Yan Li^{1,2} | Yao Yao^{5,6} | Mingyan Fang^{1,2} | Chen Bai^{6,7} | Peng An⁸ | Huashuai Chen⁹ | Zhihua Chen^{1,2} | Biyao Tang⁴ | Juan Shen¹⁰ | Xiaotong Gao⁴ | Mingrong Zhang¹⁰ | Pengyu Chen⁴ | Tao Zhang^{1,2} | Huijue Jia^{1,2} | Xiao Liu^{1,2} | Yong Hou^{1,2} | Huanming Yang^{1,2} | Jian Wang^{1,2} | Fudi Wang^{4,8} | Xun Xu^{1,2,11} | Junxia Min⁴ | Chao Nie^{1,2} | Yi Zeng^{5,6}

¹BGI-Shenzhen, Shenzhen, China

²China National Genebank, Shenzhen, China

³BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China

⁴The First Affiliated Hospital, Institute of Translational Medicine, School of Medicine, Zhejiang University, Hangzhou, China

⁵Center for the Study of Aging and Human Development, Medical School of Duke University, Durham, USA

⁶Center for Healthy Aging and Development Studies, National School of Development, Raissun Institute for Advanced Studies, Peking University, Beijing, China

⁷School of Labor and Human Resources, Renmin University, Beijing, China

⁸Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing, China

⁹Business School of Xiangtan University, Xiangtan, China

¹⁰BGI Genomics, BGI-Shenzhen, Shenzhen, China

¹¹Guangdong Provincial Key Laboratory of Genome Read and Write, Shenzhen, China

Correspondence

Xun Xu, BGI-Shenzhen, Shenzhen, China.
Email: xuxun@genomics.cn

Junxia Min, The First Affiliated Hospital, Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China.
Email: junxiamin@zju.edu.cn

Chao Nie, BGI-Shenzhen, Shenzhen, China
Email: niechao@genomics.cn

Yi Zeng, Center for Healthy Aging and Development Studies, National School of Development, Raissun Institute for Advanced Studies, Peking University, Beijing, China.
Email: zengyi@nsd.pku.edu.cn

Funding information

National Key Research and Development Program of China: 2018YFC2000400; 2020YFC2002902; National Natural

Abstract

There is growing interest in studying the genetic contributions to longevity, but limited relevant genes have been identified. In this study, we performed a genetic association study of longevity in a total of 15,651 Chinese individuals. Novel longevity loci, *BMPER* (rs17169634; $p = 7.91 \times 10^{-15}$) and *TMEM43/XPC* (rs1043943; $p = 3.59 \times 10^{-8}$), were identified in a case-control analysis of 11,045 individuals. *BRAF* (rs1267601; $p = 8.33 \times 10^{-15}$) and *BMPER* (rs17169634; $p = 1.45 \times 10^{-10}$) were significantly associated with life expectancy in 12,664 individuals who had survival status records. Additional sex-stratified analyses identified sex-specific longevity genes. Notably, sex-differential associations were identified in two linkage disequilibrium blocks in the *TOMM40/APOE* region, indicating potential differences during meiosis between males and females. Moreover, polygenic risk scores and Mendelian randomization analyses revealed that longevity was genetically causally correlated with reduced risks of multiple diseases, such as type 2 diabetes, cardiovascular diseases, and arthritis. Finally, we incorporated genetic markers, disease status, and lifestyles to classify longevity

Xiaomin Liu, Zijun Song, Yan Li and Yao Yao contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Aging Cell* published by the Anatomical Society and John Wiley & Sons Ltd.



Science Foundation of China, Grant/Award Number: 72061137004, 81903392 and 81941021; Guangdong Provincial Key Laboratory of Genome Read and Write, Grant/Award Number: 2017B030301011; Science, Technology and Innovation Commission of Shenzhen, Grant/Award Number: JCYJ20170412153100794 and JCYJ20180507183615145; the U.S. National Institute of Aging of National Institute of Health (P01AG031719)

or not-longevity groups and predict life span. Our predictive models showed good performance (AUC = 0.86 for longevity classification and explained 19.8% variance of life span) and presented a greater predictive efficiency in females than in males. Taken together, our findings not only shed light on the genetic contributions to longevity but also elucidate correlations between diseases and longevity.

KEYWORDS

centenarians, human genetics, life span, longevity, sex difference

1 | INTRODUCTION

The average human life expectancy has been rising for decades (Greene, 2001; Oeppen & Vaupel, 2002), and it was recently estimated that the number of long-lived individuals (more than 90 years old) was 63.5 million worldwide as of 2020 (United Nations, 2019). It is clear that longevity represents a complex trait that is influenced by genetic and environmental factors and their interactions (Passarino et al., 2016). Twin studies (Herskind et al., 1996; Skytthe et al., 2003) have estimated that the heritability of longevity is approximately 20%–30% in modern societies, and the proportion increases to approximately 40% for long-lived individuals (Hjelmborg et al., 2006; Perls et al., 2000; Terry et al., 2007; van den Berg et al., 2019). Although longevity is considered to exhibit relatively high heritability, limited genetic loci related to this trait have been identified in previous genome-wide association studies (GWAS; Deelen et al., 2014; Joshi et al., 2017; McDaid et al., 2017; Sebastiani et al., 2012; Zeng et al., 2016). Apolipoprotein E (APOE) is the only gene that has been replicated by multiple independent GWAS meta-analyses (Deelen et al., 2019; Joshi et al., 2017; McDaid et al., 2017). One recent meta-analysis revealed rs7676745 near *GPR78* as a novel locus (Deelen et al., 2019). In European populations, other GWAS meta-analysis studies have replicated several longevity genes, including *CHRNA3/5*, *CDKN2A/B*, *SH2B3*, and *FOXO3A* (Joshi et al., 2017; McDaid et al., 2017). Our previous GWAS in a Chinese population additionally identified *IL6* and *ANKRD20A9P* (Zeng et al., 2016). One possible reason for the lack of replication could be the variation in phenotype definitions. Some studies have compared old cases with young controls. The selection of age cutoffs varies among different studies. A recent study conducted on multi-ethnic datasets used the 90th/99th survival percentile as the age cutoff (Deelen et al., 2019). Some other studies have used more extrema cutoffs, with only centenarians being included among the cases (Sebastiani et al., 2012, 2017; Zeng et al., 2016). The use of different definitions for cases and controls may lead to heterogeneity. Continuous phenotypes have also been widely used for longevity genetic studies. A person's life span is the most obvious phenotype. Nevertheless, many ongoing cohorts contain younger participants, resulting in the limited sample size of people with exact death dates. Therefore, the parental life span has been used as an alternative phenotype (McDaid et al., 2017; Timmers et al.,

2019). The predicted life span based on family history life span has also been used (Yashin et al., 2018). However, because the lifestyles, health care, and other environmental factors are quite different between generations, the use of parental life spans might introduce substantial bias. Notably, the previous studies often focused on Western populations; very few studies have focused on Asians, which account for 42% of the long-lived population worldwide (United Nations, 2019). The Chinese Longitudinal Healthy Longevity Survey (CLHLS) encompasses a nationwide sample of long-lived adults and the “young old” in China, with up to 20 years of follow-up which enable us to track longevity and mortality. Therefore, CLHLS can provide an ideal dataset in analyzing the association of genetic and non-genetic data with life span in humans.

In addition to studying the genetics of longevity and life span, age-related diseases and their correlations with longevity have attracted much attention (Sakaue et al., 2020). In either human centenarians or long-lived animals, it has long been observed that longevity and the occurrence of diseases, such as cardiovascular and cerebral stroke, are inversely correlated either genetically or experimentally (Altmann-Schneider et al., 2013; Hammond et al., 1971; Rosa et al., 2019; van der Lee et al., 2019). A previous study, using genetic data of parental life span, reported genetic correlations between several complex traits and mortality in a general population of European ancestry (Joshi et al., 2017). Our study defined individuals with ages greater than 90 as the longevity group, rather than the parental survival which is a debatable longevity phenotype. Therefore, a systematic exploration of the correlation between longevity and complex diseases in the current study may reveal more information.

Another research interest is to predict longevity and life span based on age-related diseases and genetic markers. The polygenic risk score (PRS) generated from the summary statistics of association studies is a commonly used predictor for genetic factors. For example, a recent genetic study reported that a polygenic score could identify people with the top 10% parental survival PRS, who might outlive an average of 5 years those with the bottom 10% parental survival PRS (Timmers et al., 2019). In addition, circulating glucuronic acid levels (Ho et al., 2019) or telomeres (Whittemore et al., 2019) have been used as biomarkers for life span prediction. To date, there are very few studies that explore the potential of life span prediction by using a combination of genetic data, disease conditions, and lifestyle factors.



Here, we performed a large-scale integrated analysis based on 15,651 Chinese individuals from CLHLS to identify the longevity genes, to explore the relationships between diseases and longevity, and to apply these longevity-related factors for life span prediction. This study including 2,509 centenarians is one of the largest centenarian studies in the world (Deelen et al., 2019; Sebastiani et al., 2017; Timmers et al., 2019). We firstly designed a customized SNP chip using a carefully selected set of SNPs that captured 27,656 candidate variants correlated with longevity, age-related diseases, and immunity. Next, we carried out a candidate-gene association analysis on the age-stratified phenotype ("cases" were defined as individuals surviving 90 years or over, while the "controls" had an age of less than 75, which is the average life span in China) and life span, respectively. Then, we performed a meta-analysis incorporating the current dataset and our previously published GWAS dataset by removing the overlapped samples. Moreover, we evaluated the polygenic prediction of diseases on longevity using polygenic risk score (PRS) analysis and inferred causal relationships between longevity and diseases using the bi-directional Mendelian randomization method. Finally, we built predictive models for longevity, and life span by integrating genetic factors, disease status, and lifestyles. Overall, this study aimed to reveal the sex-combined and sex-specific longevity genes/pathways and investigate their predictive effectiveness on longevity and life span.

2 | MATERIALS AND METHODS

2.1 | Participants and phenotypes

This study included a total of 15,651 individuals from the Chinese Longitudinal Healthy Longevity Surveys (CLHLS), which were conducted in 1998, 2000, 2002, 2005, 2008, 2011, and 2014 in a randomly selected half of the counties and cities in 22 out of 31 provinces in China. The primary dataset (dataset 1) included 13,228 individuals with ages ranging from 30 to 114. All individuals were genotyped by using a well-designed customized chip targeting approximately 27 K longevity-related SNPs. These candidate SNPs were selected based on previously published associations with longevity, chronic diseases, and health indicators. For replication purposes, a dataset 2 included 4477 individuals based on our previous study (Zeng et al., 2016). 2054 samples were overlapped between the two datasets.

Demographic and clinical information (i.e., diseases) was recorded for participants in this study. Phenotypic data were collected using internationally standardized questionnaires adapted to the Chinese cultural and social context. The CLHLS study was approved by the Biomedical Ethics Committee of Peking University (IRB00001052-13074). All participants or their legal representatives signed written consent forms in the baseline and follow-up surveys.

2.2 | Customized SNP chip design

We customize a SNP chip containing 27,656 selected longevity and disease-related SNPs for targeted genotyping (Table S1). The selected SNPs could be characterized as corresponding to five major components (Tables S2 and S3): (1) 11,893 SNPs associated with longevity based on our previous CLHLS GWAS study on 4477 Chinese individuals (Zeng et al., 2016); (2) 1881 reported longevity SNPs based on the other previously published longevity studies, including the European Union (EU) longevity (Deelen et al., 2014) & New England centenarian study (NECS; Sebastiani et al., 2012), the Long Life Family Study (LLFS; Bae et al., 2013), and the Framingham Heart Study (FHS; Lunetta et al., 2007); (3) 3966 reported SNPs associated with diseases in the NHGRI-EBI GWAS catalog (MacArthur et al., 2017); (4) 7260 SNPs associated with health indicators in the CLHLS and the Health and Retirement Study (HRS; Tanaka et al., 2017); and (5) 2656 tagging SNPs for imputing the alleles in the human major histocompatibility complex (MHC) region.

2.3 | Sample filtering

The samples were required to meet 3 selection criteria: (1) a genotype calling rate >90%; (2) no existing population stratification according to a multidimensional scaling (MDS) procedure implemented in PLINK v1.07, based on which individuals deviating from the main population cluster were removed; and (3) no inclusion of duplicates or first-degree relatives when evaluating pairwise through identity by descent (IBD). After sample filtering, 12,664 samples were included in the dataset 1.

2.4 | Variant filtering

To determine the high-quality genotypes, we applied a conservative inclusion threshold for variants: (1) minor allele frequency >5%, (2) genotype calling rate >90%; and (3) Hardy-Weinberg equilibrium (HWE) $p > 10^{-5}$. To further confirm the quality of the genotypes, we calculated the concordance rate of the genotypes using 2,054 samples that overlapped between dataset 1 and dataset 2. Then, we removed the variants with a concordance rate <0.9 (Figure S7), which largely eliminated the bias caused by two different arrays (Illumina ZhongHua and Affix arrays). After variant filtering, 23,769 out of the 27,656 variants remained in dataset 1, and 818 K out of the 900 K variants remained in dataset 2.

2.5 | Imputation

We performed imputation analysis by pre-phasing genotypes with SHAPEIT v2.5 (Delaneau et al., 2011), and then imputing variants from the 1000 Genomes Project released on October 2014 with



2504 samples (<http://1000genomes.org>) as a reference panel using IMPUTE2 v2.3.1 (Howie et al., 2009). SNPs with a quality score (R^2) >0.9 were included after imputation. After further quality control filtering for SNPs as described above, we eventually obtained 287 K SNPs from 12,664 individuals in dataset 1 and 5.6 M SNPs from 4,477 individuals in dataset 2 for the subsequent genetic association analyses.

2.6 | MHC analysis

To identify potential MHC associations for longevity, 2,656 MHC tag SNPs were included in the 27 K arrays for dataset 1. Then, we used beagle 5 (Browning et al., 2018) with the HAN-MHC datasets as a reference panel to impute MHC alleles, and the imputation accuracy was 0.96 at the two-digit level as previously described (Zhou et al., 2016). In dataset 2, the samples were genotyped using Illumina HumanOmniZhongHua-8 BeadChips tagging 900,015 SNPs, among which 8,350 SNPs were located in the MHC region. We imputed the MHC alleles using the same procedure applied for dataset 1 and obtained 104 imputed HLA alleles presented in both two datasets. For each dataset, 104 tests were performed in the cases and controls. In each test, one allele was compared with the other 103 alleles grouped together. The allelic 2×2 contingency table for a specific HLA allele contained the counts of that allele and the counts of the other 103 alleles in cases and controls. We next performed a meta-analysis of the two datasets for the 104 imputed HLA alleles for longevity. Finally, a Bonferroni-corrected $p < 0.0005 = 0.05/104$ for 104 alleles was defined as significant.

2.7 | Association analysis for longevity

We performed genetic association analysis of 287 K imputed SNPs in dataset 1. More specifically, after sample filtering, a total of 8,490 individuals (4,662 cases with an age ≥ 90 and 3,828 controls with an age <75 ; 75 is the average life span of Chinese individuals) in dataset 1 were used for a case-control association analysis. We then performed association analysis in dataset 2 (Zeng et al., 2016). Since 1,922 individuals in dataset 2 were overlapped with 8,490 case/control samples in dataset 1, a case-control association analysis was performed in 2,555 independent samples by removing the 1,922 overlapped samples from 4,477 samples of dataset 2. The 2,555 independent samples included 1,105 centenarians' cases and 1,450 controls with age <65 . For each dataset, we applied logistic regression to calculate the p -values and odds ratio (ORs) of the SNPs by adjusting for sex and the top two MDS dimensions using PLINK 1.07. Next, a meta-analysis was performed on the two case-control association results, using inverse-variance weighted fixed-effect meta-analysis in METAL software (<https://genome.sph.umich.edu/wiki/METAL>). To further replicate the results, we also reviewed association results from previous literatures, including EU & NECS, LLFS & FHS longevity GWAS, and studies in the GWAS catalog.

To investigate the correlations between the identified longevity-related SNPs and diseases and the other traits, we reviewed diseases GWAS in this study (see Section 2.12 below). Then, we downloaded the summary statistics data from the Japan BioBank, a study of 300,000 Japanese citizens suffering from cancers, diabetes, rheumatoid arthritis, and other common diseases (Triendl, 2003). Similarly, we searched the longevity SNPs in the summary statistics data from Japan Biobank to examine their associations with metabolic traits and diseases.

2.8 | Sex-specific association analysis for longevity

We performed sex-specific genetic association analyses in males and females separately. Male-specific variants were identified as those that (1) were significantly associated with longevity in males ($p_{\text{male}} < 5 \times 10^{-8}$) but not significant in females ($p_{\text{female}} > 0.05$), and (2) exhibited a nominally significant sex difference (p -value testing for difference in sex-specific effect estimates, $p_{\text{difference}} < 0.05$; formula 1). Female-specific variants were identified as (1) significantly associated with longevity in females ($p_{\text{female}} < 5 \times 10^{-8}$) but not significant in males ($p_{\text{male}} > 0.05$), and (2) exhibited a nominally significant sex difference ($p_{\text{difference}} < 0.05$; formula 1). For each variant, we calculated $p_{\text{difference}}$ testing for the difference between the male-specific and female-specific beta estimates β_m and β_f using the T-statistic (formula (1))

$$T = \frac{(\beta_m - \beta_f)}{\sqrt{SE_m^2 + SE_f^2 - 2 * \text{cor}(\beta_m, \beta_f) * SE_m * SE_f}} \quad (1)$$

where SE_m and SE_f are the standard errors of the beta estimations in different sex groups.

2.9 | Functional annotation and enrichment analysis

The significant loci with $p < 10^{-5}$ identified in the sex-combined genetic association analysis with longevity were mapped to genes using SNP2GENE in FUMA (Watanabe et al., 2017; <http://fuma.ctglab.nl/>). We mapped variants to genes based on physical distance within a 20-kb window by the positional mapping method. The mapped genes were further investigated using the GENE2FUNC procedure, which provides hypergeometric tests for the list of enriched mapping genes in 53 GTEx tissue-specific gene expression sets, 7,246 MSigDB gene sets, and 2,195 GWAS catalog gene sets. Using the GENE2FUNC procedure, we examined whether the mapped genes were enriched in specific diseases or traits in the GWAS catalog as well as whether enriched in specific Gene Ontology (GO) and Kyoto Encyclopedia of Genes (KEGG) categories. Significant results were selected if a false discovery rate (FDR)-corrected $p < 0.05$ was observed.



To identify sex-specific longevity pathways, the best-fit p -value cutoffs of 5×10^{-5} and 0.0015 (calculated by using PRSice; Euesden et al., 2015, software with the BEST FIT command) were used to select SNPs in females and males, respectively. SNPs were annotated and enriched for the pathways obtained from MSigDB version 5.2 limited to the KEGG, Reactome, and GO databases. A SNP was annotated to a gene if it fell within the interval of the coding sequence with the upstream and downstream 50 kb flanking regions. We performed the gene set enrichment analysis using MAGMA (de Leeuw et al., 2015) and sex-specific pathway gene sets, and those showing a FDR-corrected $p < 0.05$ were regarded as significant.

2.10 | Observational correlation analysis

We had detailed questionnaire information including sex, age, diseases, cognition, and lifestyle factors. The observational correlation analysis was performed to assess the statistical relationship (i.e., the correlation) between longevity and these influencing factors, which were evaluated by multivariable linear regression analysis while adjusted for sex and the top two MDS (Table S13). In the multivariable linear regression model, the phenotype is longevity trait with longevity cases as 1 and middle-aged controls as 0. The variables included diseases, cognition, and lifestyle factors, for example, suffering from a disease or not was defined as 1 and 0, respectively.

2.11 | Polygenic risk Scores (PRS) analysis

In this cohort study, we calculated weighted polygenic risk scores based on 3,966 known susceptibility markers from the GWAS catalog for many age-related diseases (Table S5). We imputed the missing risk alleles and corresponding beta weights whenever possible by checking the details in the original reports. Markers were coded additively, and the logarithms of the reported odds ratios were used as weights. All markers were clumped by pairwise linkage disequilibrium ($r^2 > 0.8$) prior to constructing the polygenic risk score. Each disease containing at least five SNPs was used to generate the PRS for each individual, and 87 diseases were ultimately included (Table S14). PRS analysis was performed to calculate the correlations of longevity and disease risks not only in all individuals but also in males and females, respectively.

2.12 | Mendelian randomization (MR) analysis

We had detailed disease records for each individual. Therefore, we performed the GWAS for each of the disease types using the same approach that we applied for the longevity association analysis and used summary statistics data for longevity and various diseases for MR analysis. To calculate the causal effect of longevity on diseases, as well as the effects of diseases on longevity, we performed a bidirectional MR analysis using four different MR methods, including the

GSMR method (Zhu et al., 2018) in GCTA tool and inverse-variance weighting (IVW), weighted median and MR-Egger regression methods implemented in the “TwoSampleMR” R package for robust validation. A consistent effect across the four methods is less likely to be a false positive. In the process of the MR analysis, we selected independent SNPs as instrumental variables, setting a linkage disequilibrium threshold of $r^2 < 0.2$ in a 500-kb window. We explored multiple settings for instrument strength with $p < 10^{-3}$, $p < 10^{-4}$, and $p < 10^{-5}$, respectively. We used the MR-Egger intercept implemented in MR-Egger regression to test for the presence of directional pleiotropy.

2.13 | Survival analysis

By the last interview, 3,040 of the 12,664 individuals were reported to have died by their families. To study the relationship between genotypes and life span, we used age and live/dead status as phenotypes, and we used a multivariate Cox proportional regression model to perform association analysis. The model was implemented with the “coxph” function from the survival package in R 3.5.1. In the model, the individuals were either dead (status 1), alive (status 0); or missing at the follow-up interview were subjected to censoring. The surviving subjects were calculated according to age and censoring parameters. Then, the Cox regression was performed using all genotypes and sex as independent variables and the surviving status as the dependent variable. The survival curves were plotted using the “survfit” function in R 3.5.1.

2.14 | Lasso regression prediction

The prediction analysis was completely independent from the association analyses, and all the 23,769 SNPs incorporated with 19 disease statuses, five lifestyle measurements, and sex were entered into the least absolute shrinkage and selection operator (Lasso). Lasso is a supervised machine learning method that can select a subset of SNPs to achieve the best prediction efficiency. The disease status, lifestyle measurements, and SNP genotypes were imputed separately. Disease status and lifestyles were imputed using the MICE package in R 3.5.1. All the SNPs remaining after QC ($n = 23,769$) were imputed internally without using any reference panel, because the reference panel-based imputation leverages linkage disequilibrium information (e.g., highly correlated SNPs will be imputed), which is redundant information in terms of prediction. The missing genotypes were imputed using Beagle 5.0 (Browning et al., 2018). Together with sex, 23,824 predictors were entered into the Lasso regression (Tibshirani, 1996). The whole dataset was split into 80% and 20% subsets for model training and testing. Fivefold cross-validation was conducted for the training dataset. The training process for Lasso regression included feature selection and model fitting. Only one of the features with redundant information was selected for modeling, such as SNPs in high linkage disequilibrium. For

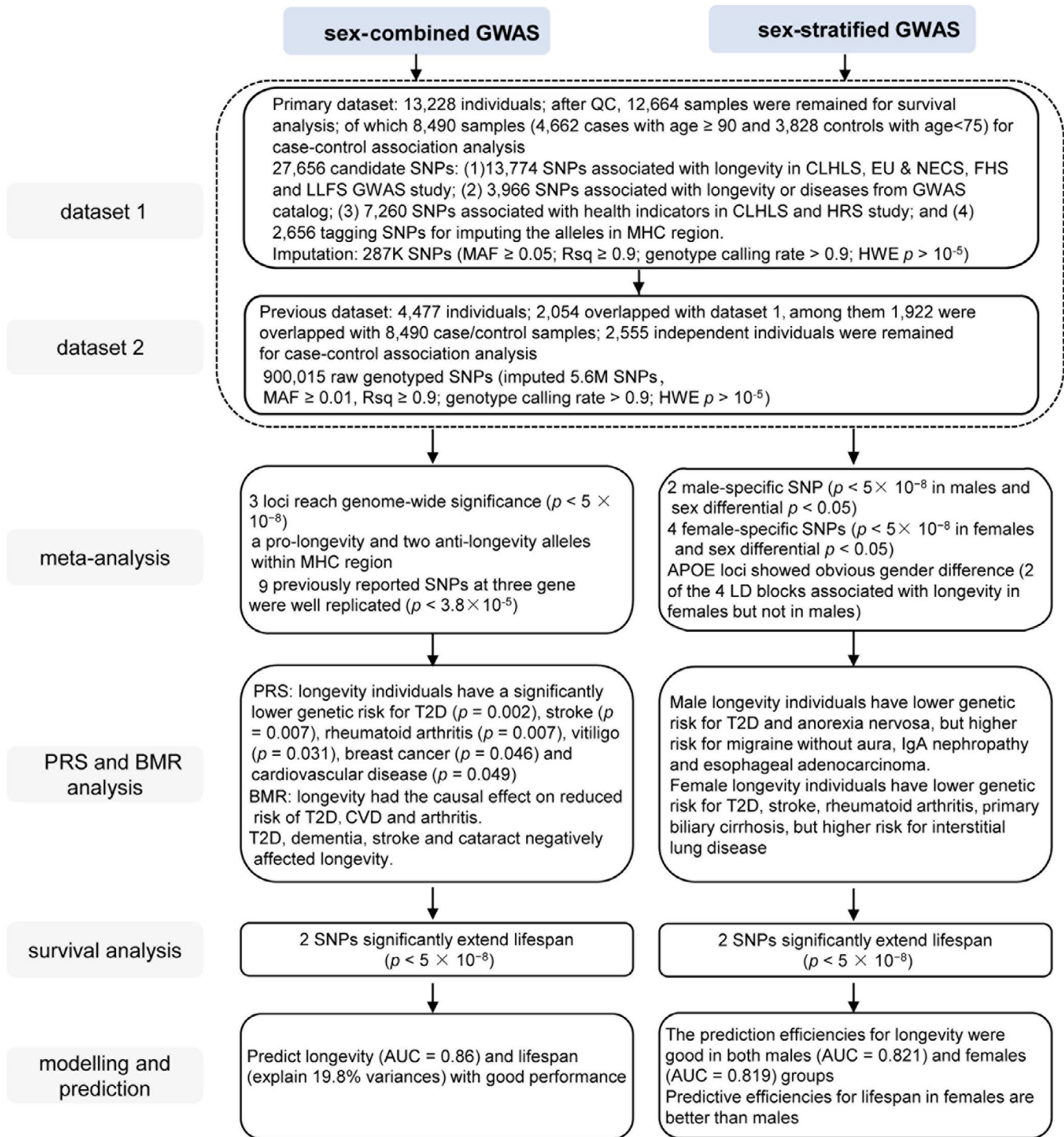


FIGURE 1 Study design and workflow. To investigate the longevity-associated genes/pathways, we first carefully selected and designed a customized SNP chip that captured 27,656 candidate variants mainly for longevity as well as disease, health indicator, and immunity. Then, we genotyped these SNPs in a large sample of 13,228 individuals. Next, we carried out the genetic association analyses using age-stratified phenotype ("cases" were defined as individuals surviving past 90 years of age and "controls" with age less than average life span of 75) as well as incorporating all individuals' age and the survival status as phenotype, respectively. Furthermore, we performed meta-analysis together with previous dataset (removing overlapped samples) to identify the longevity genes in gender-combined and gender-stratified groups, respectively. In addition, we evaluated polygenic prediction of diseases on longevity using polygenic risk score (PRS) analysis and inferred causal relationships between longevity and diseases using bidirectional Mendelian randomization (MR) method. Finally, we built the prediction model for longevity and life span using all existing factors.



TABLE 1 Three loci associated with longevity at genome-wide significance

SNP	CHR	BP	Gene	EA	NEA	Dataset 1			Dataset 2 (removing overlapped samples)			Meta-analysis			Associated diseases
						β	SE	p	β	SE	p	β	SE	p	
rs17169634	7	34093997	BMPER	G	A	0.370	0.043	3.87E-18	-0.097	0.109	0.376	0.308	0.040	7.91E-15	Arthritis ($p = 3.76E-06$); prostate cancer ($p = 6.32E-03$)
rs2075650	19	45395619	TOMM40	G	A	-0.261	0.052	6.76E-07	-0.442	0.111	7.26E-05	-0.293	0.047	6.17E-10	Dementia ($p = 2.40E-04$); arthritis ($p = 3.01E-03$)
rs1043943	3	14183410	TMEM43	T	C	0.165	0.033	4.24E-07	0.081	0.042	0.016	0.146	0.029	3.59E-08	Parkinson's disease ($p = 0.045$)

Note: After standard quality control process, 8,490 individuals (4,662 cases with an age ≥ 90 and 3,828 controls with an age < 75 ; 75 is the average life span of Chinese individuals) in dataset 1 and 2,555 independent individuals (removing 1,922 samples overlapped between 8,490 individuals in dataset 1 and 4,477 individuals in dataset 2; remaining 1,105 centenarians and 1,450 controls with age < 65) in dataset 2 were included for case-control association analysis. A meta-analysis of the two datasets was performed. The BP (base-pair position) is based on Genome Reference Consortium Human Build 37 (GRCh37).

Abbreviations: EA, effect allele; NEA, not effect allele; SE, standard error.

longevity prediction, AUCs were calculated to evaluate prediction efficiency. For the life span prediction, the explained variance for life span was estimated using linear model. The predictions were also performed in the male and female groups separately.

3 | RESULTS

3.1 | Study subjects and design

This study was composed of two datasets including a total of 15,651 individuals. The first included 13,228 individuals with 27,656 longevity- and disease-related SNPs genotyped by using our customized SNP chip (Tables S1–S3, see also Section 2). After implementing the standard quality control procedures, 12,664 samples and 23,769 out of the 27,656 SNPs remained for subsequent analysis (Figure S1A). Reference-based imputation enlarged the dataset into 287,000 SNPs. The second dataset was the GWAS set that we previously published, including 4,477 samples (2,178 centenarians and 2,299 middle-aged controls, Figure S1B) and 5.6 M imputed SNPs (Zeng et al., 2016). The two datasets included 2,054 overlapping individuals, and the genotype concordances of SNPs for the same individual were measured for quality control. The discordant SNPs (genotype concordance < 0.9) were removed, and the remaining SNPs were imputed for association analysis. The analysis flow is presented in Figure 1.

We first performed a case-control association analysis on 8,490 individuals (4,662 cases with age ≥ 90 and 3,828 controls with age < 75 , Figure S1A) from dataset 1. Twelve SNPs achieved significance after *Bonferroni's* correction (Table S4; $p < 1.81 \times 10^{-6} = 0.05/27,656$). In dataset 2, 1,922 individuals, who were overlapped with the 8,490 case/control samples of dataset 1, were removed. We then investigated the significance of these 12 SNPs in the independent dataset 2 with 2,555 individuals. Two SNPs were also nominally significant in the same direction (Table S4; $p < 0.05$). Therefore, we performed both sex-combined and sex-stratified meta-analyses of these two independent datasets including 11,045 individuals to further identify potential longevity genes and pathways. Meanwhile, we carried out survival analysis, polygenic risk score (PRS) prediction, and Mendelian randomization (MR) in all 12,664 individuals from dataset 1. The survival analysis utilized age and the live/dead status as a phenotype to identify SNPs associated with life span. Furthermore, we used PRS analysis and bidirectional MR method to infer causal relationships between longevity and diseases. Finally, we built the prediction model for longevity and life span in 7,807 individuals from dataset 1 with disease status, lifestyle records, and SNP genotypes.

3.2 | Novel longevity genes revealed by meta-analysis

As shown in Figure 1, we identified three loci that were significantly associated with longevity ($p < 5 \times 10^{-8}$) in the meta-analysis of the two

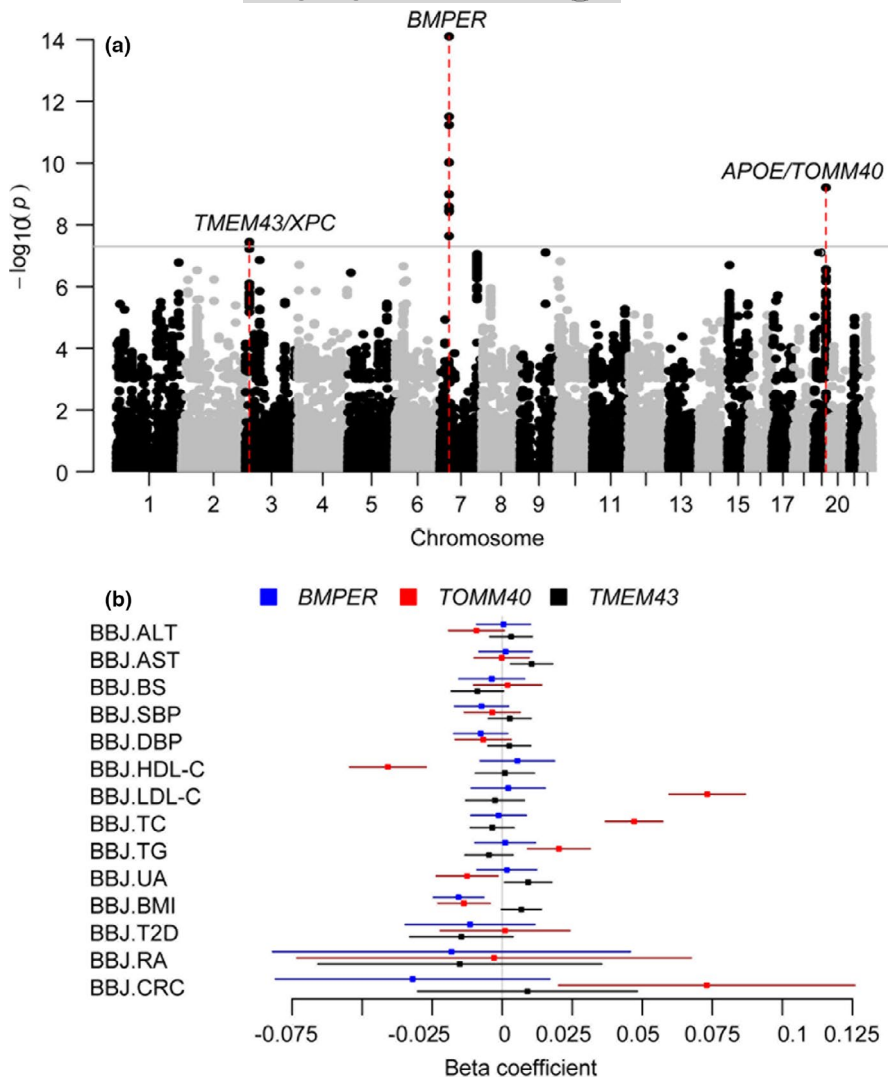


FIGURE 2 Identification of novel longevity genes by meta-analysis. (a) Manhattan plot showing the association results by meta-analysis for longevity. The red line represents a genome-wide significant p -value (5×10^{-8}). The gray lines showed the top three loci associated with longevity, including rs17169634 ($p = 7.91 \times 10^{-15}$) at gene *BMPER*, rs2075650 ($p = 6.17 \times 10^{-10}$) at *TOMM40/APOE* loci, and rs1043943 ($p = 3.59 \times 10^{-8}$) at 3'-UTR of *TMEM43*. (b) Crossing references with GWAS studies from BioBank Japan (BBJ) showed that the three longevity SNPs also linked to multiple traits. ALT, alanine aminotransferase; and CRC, colorectal cancer; AST, aspartate aminotransferase; BMI, body mass index; BS, blood sugar; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; RA, rheumatoid arthritis; SBP, systolic blood pressure; T2D, type 2 diabetes; TC, total cholesterol; TG, total triglyceride; UA, uric acid.

datasets. Among these 3 identified longevity loci, one is the well-known locus located near the *TOMM40/APOE* region. The two newly identified genes/loci include *BMPER* and *TMEM43* (Table 1; Figure 2A; Figure S2). The top signal, rs17169634 ($p = 7.91 \times 10^{-15}$), is located in the intronic region of the *BMPER* gene and has been reported to be associated with Alzheimer's disease (Nelson et al., 2014; Table S5). The second significant signal is the well-known longevity locus *TOMM40/APOE* (rs2075650; $p = 6.17 \times 10^{-10}$). The third top SNP, rs1043943 ($p = 3.59 \times 10^{-8}$), is located in the 3'-UTR of *TMEM43*, which is in strong linkage disequilibrium (LD) with rs2228001 ($r^2 = 0.95$; $p = 1.13 \times 10^{-7}$), a missense mutation in the *XPC* gene. It has been reported that rs1043943 might regulate the expression of *XPC*, which is a nucleotide excision repair (NER) gene involved in DNA damage repair, and the deletion of *XPC* leads to the development of lung tumors in mice (Hollander et al., 2005). We investigated these three identified longevity-associated SNPs in the two largest relevant meta-analyses results (Deelen et al., 2019; Timmers et al., 2019; Table S6). The *TOMM40/APOE* (rs2075650) locus was significant in all three studies and shows consistent effect directions for the minor allele. It is also the most notable genetic loci across many GWAS in multiple ethnic populations. The other two loci were not replicated. Rs1043943 in

TMEM43 was nearly nominal significant ($p = 0.057$) in parents' life span study by Timmers et al. but not significant in Deelen's study ($p > 0.5$). While for rs17169634 in *BMPER*, although Deelen's study showed nominally significance ($p = 0.033$), the effect of minor allele G was in the opposite direction with our results. We further performed association analyses for diseases in dataset 1 and found that the *BMPER* locus was associated with arthritis ($p = 3.76 \times 10^{-6}$) and prostate cancer ($p = 6.32 \times 10^{-3}$), and the G allele for SNP rs17169634 has decreased effects on the risk of arthritis and prostate cancer. *TOMM40/APOE* locus linked to dementia ($p = 2.40 \times 10^{-4}$) and arthritis ($p = 3.01 \times 10^{-3}$), and *TMEM43* was associated with Parkinson's disease ($p = 0.045$). Interestingly, our three longevity SNPs have also been linked to multiple metabolic traits in GWASs on the Japan BioBank dataset (Triendl, 2003; Figure 2B). Specifically, *BMPER* is associated with body mass index (BMI); *TOMM40/APOE* is associated with low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), total cholesterol (TC), total triglyceride (TG), and colorectal cancer (CRC), and *TMEM43* is associated with aspartate aminotransferase (AST) and blood uric acid (UA).

Since we enriched SNPs located in the MHC region, we next imputed 104 HLA alleles using the HAN-MHC reference panel (Zhou



TABLE 2 Sex-stratification results for longevity

SNP	Chr.	position	EA	NEA	Gene	Males						Females						Sex difference	
						Dataset 1		Dataset 2		Meta-analysis		Dataset 1		Dataset 2		Meta-analysis		β	p
						β	p	β	p	β	p	β	p	β	p	β	p		
Six sex-specific loci associated with longevity at genome-wide significant levels																			
Two male-specific loci																			
rs2308910	6	33037587	a	t	HLA-DPA1	0.878	3.28E-09	0.107	0.8108	0.802	1.27E-08	0.141	0.254	0.045	0.8488	0.120	0.2711	1.35E-04	
rs16981095	19	16190863	g	t	TPM4	0.455	1.66E-09	-0.093	0.6584	0.392	3.29E-08	0.184	8.38E-03	0.063	0.6117	0.155	0.0109	1.09E-02	
Four female-specific loci																			
rs10490092	2	58892049	c	t	FLJ30838	0.123	0.131	0.003	0.988	0.110	0.1537	0.360	2.43E-06	0.364	6.20E-03	0.361	4.01E-08	1.31E-02	
rs6967652	7	140576149	g	a	BRAF	0.163	0.0181	-0.366	0.0832	0.112	0.0876	0.451	3.04E-11	-0.101	0.4049	0.320	3.91E-08	1.88E-02	
rs2075650	19	45395619	g	a	TOMM40	-0.168	0.0305	-0.467	0.04906	-0.197	7.64E-03	-0.336	1.87E-06	-0.436	5.64E-04	-0.360	5.12E-09	4.87E-02	
rs73329134	7	34071880	t	c	BMPER	0.207	7.61E-03	-0.050	0.812	0.176	0.0155	0.499	9.76E-12	-0.078	0.5462	0.359	1.82E-08	3.92E-02	
Sex difference results for the three significant loci identified in sex-combined association																			
rs17169634	7	34093997	g	a	BMPER	0.288	3.36E-06	-0.122	0.5593	0.255	1.79E-05	0.444	7.25E-14	-0.093	0.472	0.350	8.17E-11	2.34E-01	
rs2075650	19	45395619	g	a	TOMM40	-0.168	0.0305	-0.467	0.0491	-0.197	7.64E-03	-0.336	1.87E-06	-0.436	5.64E-04	-0.360	5.12E-09	4.87E-02	
rs1043943	3	14183410	t	c	TMEM43	0.128	8.27E-03	0.069	0.5636	0.100	0.0254	0.197	9.42E-06	0.137	0.05323	0.180	1.75E-06	1.76E-01	

Note: Male-specific analysis was performed in 4,413 males (3,758 males in dataset 1 and 655 independent males in dataset 2). Female-specific analysis was performed in 6,632 females (4,732 females in dataset 1 and 1,900 independent females in dataset 2). EA, effect allele; NEA, not effect allele; the sex-differential p-values tested for difference between the male-specific and female-specific beta estimates using the T-statistic.

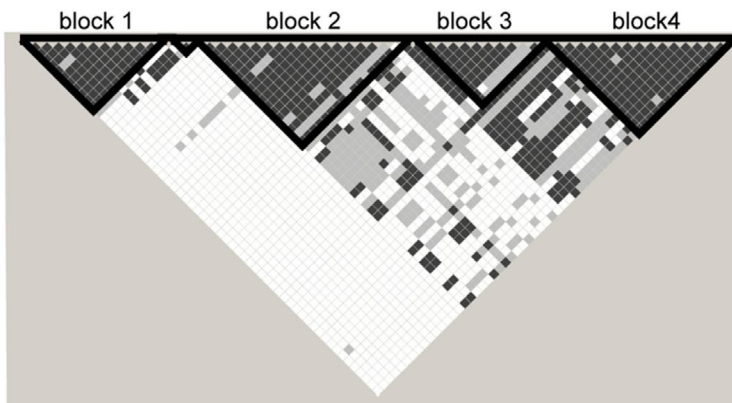
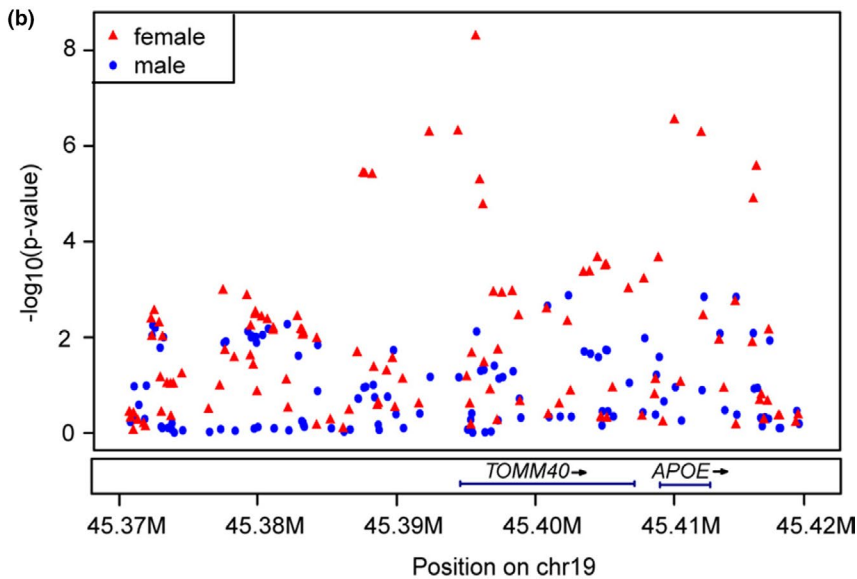
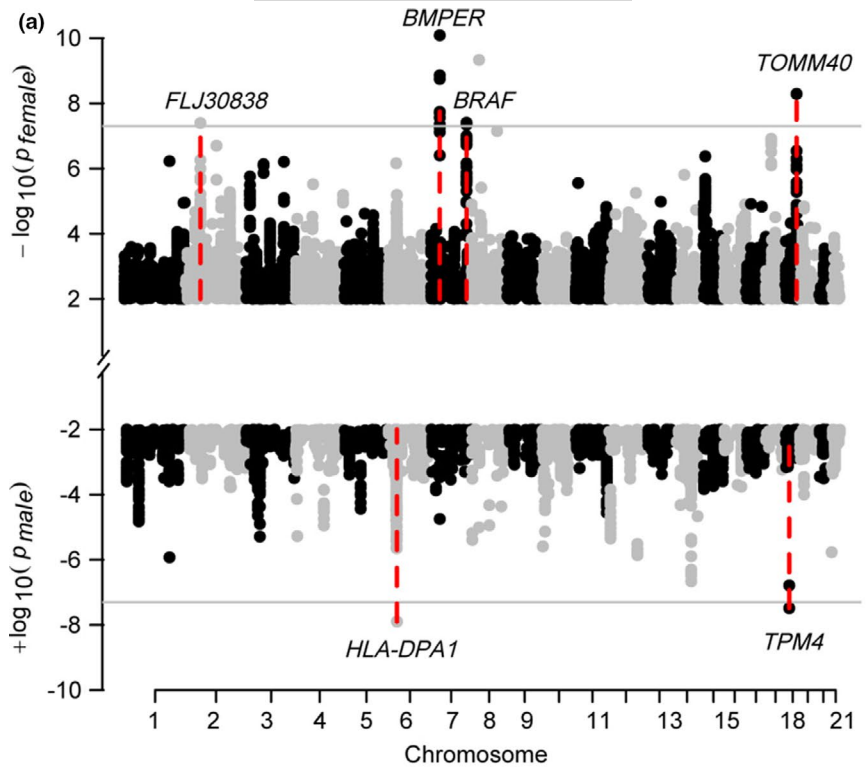


FIGURE 3 Sex differences in genetic associations with longevity. (a) Manhattan plot showing the male-specific and female-specific associations with longevity. The gray line represents the genome-wide p threshold (5×10^{-8}). Four female-specific (*FLJ30838*; *BRAF*; *BMPER*; and *TOMM40*) and two male-specific genes (*HLA-DPA1* and *TPM4*) were marked in red lines. (b) Regional plot showing the association results for *TOMM40/APOE* locus in males and females, respectively. The LD blocks of this locus were also plotted using Haploview tool, SNPs in block 2 (rs12972156, rs12972970, rs34342646, rs6857, rs71352238, rs2075650, rs34404554) and block 4 (rs769449, rs429358) showed sex differences ($p_{\text{difference}} < 0.05$), whereas SNPs in the other two blocks were not.



TABLE 3 Observational correlation and bidirectional MR analysis for longevity and diseases

Exposure	Outcome	Observational correlation			Bidirectional MR			MR pleiotropy test (Egger intercept)		
		Estimate	SE	p	Estimate	SE	p	Egger intercept	SE	p
<i>p</i> < 10 ⁻³										
Longevity	Cardiovascular disease	-0.517	0.099	1.58E-07	-0.1337	0.0234	1.16E-08	0.0063	0.0113	0.5790
Longevity	Type 2 diabetes	-1.252	0.179	2.54E-12	-0.1918	0.0403	1.91E-06	-0.0118	0.0183	0.5196
Longevity	Arthritis	-0.151	0.073	3.95E-02	-0.0681	0.0173	8.22E-05	0.0037	0.0098	0.2079
Longevity	Hypertension	-0.317	0.069	4.34E-06	-0.0485	0.0165	3.29E-03	0.0045	0.0073	0.5387
Type 2 diabetes	Longevity	-1.252	0.179	2.54E-12	-0.0307	0.0121	1.11E-02	-0.0217	0.0345	0.5347
<i>p</i> < 10 ⁻⁴										
Longevity	Cardiovascular disease	-0.517	0.099	1.58E-07	-0.1261	0.0366	5.69E-04	0.0302	0.0208	0.1473
Longevity	Type 2 diabetes	-1.252	0.179	2.54E-12	-0.1838	0.0628	3.40E-03	-0.0153	0.0318	0.6310
<i>p</i> < 10 ⁻⁵										
Longevity	Cardiovascular disease	-0.517	0.099	1.58E-07	-0.1108	0.0551	1.45E-02	0.0487	0.0345	0.1649
Longevity	Type 2 diabetes	-1.252	0.179	2.54E-12	-0.1546	0.0930	1.66E-02	-0.1059	0.0563	0.0662
Longevity	Arthritis	-0.151	0.073	3.95E-02	-0.1538	0.0408	1.63E-04	0.0818	0.0637	0.2057

Note: This table listed significant causal relationships through both observational correlation and bidirectional MR analysis (in the same effect direction and *p* < 0.05 in both two analysis) under different P thresholds. The observational relationship between longevity and diseases was evaluated by multivariable linear regression analysis while adjusted for age and top two principle components. The bidirectional MR in the table was calculated by TCGA-GSMR method and the directional pleiotropy was evaluated using the MR-Egger intercept implemented in MR-Egger regression. The details are shown in Table S15.



et al., 2016) and explored their relationships with the longevity phenotype. No SNPs in the HLA region reached genome-wide significance. While considering HLA alleles, three were observed to be significant after adjusting for the total number of reconstructed HLA alleles ($p < 0.0005 = 0.05/104$, Table S7). Among these three newly identified HLA alleles, HLA-DQB1*0609 was identified as a pro-longevity allele ($\beta = 0.42$; $p = 3.34 \times 10^{-4}$), while HLA-C*12:02 and HLA-B*52:01 were found to be anti-longevity ($\beta < 0$). HLA-C*12:02 is a susceptibility factor in the late-onset type of psoriasis in Japanese (Mabuchi et al., 2014), while HLA-B*52:01 has been reported to increase the risks of pulmonary infarction, ischemic heart disease, aortic regurgitation, hypertension, renal artery stenosis, cerebrovascular disease, and visual disturbance (Kitamura et al., 1998).

To further explore the biological functions of these identified longevity-associated signals, we investigated 300 SNPs associated with longevity at a suggestive significance level ($p < 10^{-5}$) using the FUMA tool (Watanabe et al., 2017). Interestingly, we found that these longevity-associated genes presented significantly up-regulated expression patterns in multiple specific cerebral regions, including the brain substantia nigra and brain amygdala (Figure S3). Through cross-referencing with the GWAS catalog, we found that longevity-associated SNPs showed significant enrichment related to the cerebrospinal total tau (T-tau) levels, cerebrospinal phosphorylated tau (P-tau 181p) levels, cerebral amyloid deposition (PET imaging), and various age-related diseases, such as Alzheimer's disease, age-related macular degeneration, type 2 diabetes, and ischemic stroke (Figure S4). Based on KEGG analysis, four enriched pathways included type 2 diabetes mellitus, Alzheimer's disease, and the two previously identified calcium and MAPK signaling pathways (Figure S5). Taken together, we found that longevity-related genes (*TOMM40*, *APOE*, *APOC1*, *PVRL2*, *BRAF*, *MRPS33*, *RYR2*, *BMPER*, *PMF1*, *BGLAP*, *MEIS1*, *CACNA1D*, *ZNRF3*, *CTNND2*, *CEL2F*, *KREMEN1*, *ANKRD30A*, *WFS1*, *PPP2R2C*, *GLG1*, *KLHL29*, *WSCD1*, *FUT10*, *TMEM43*) are also implicated in diseases and pathways involved in cardiovascular homeostasis, immunity, inflammation, lipid metabolism, and cognitive function. These associated gene sets are usually clustered with each other, suggesting the anti-correlation between these age-related diseases and longevity.

3.3 | Sex-specific genes associated with longevity

Regarding the sex difference in genetics of longevity, we performed a sex-stratified genetic association analysis. We compared the three genome-wide significant SNPs between sexes and found that only the *APOE/TOMM40* SNP showed a sex difference (Table 2, $p_{\text{male}} = 0.008$; $p_{\text{female}} = 5.12 \times 10^{-9}$; $p_{\text{sex difference}} = 0.049$). The other two SNPs were not significantly different between males and females ($p < 0.05$ in both males and females, and $p_{\text{difference}} > 0.05$). Notably, we identified two male-specific associations (Table 2; Figure 3A; $p < 5 \times 10^{-8}$ in males and $p_{\text{sex difference}} < 0.05$) and four female-specific associations ($p < 5 \times 10^{-8}$ in females and $p_{\text{sex difference}} < 0.05$) linked to longevity. The male-specific longevity locus, rs2308910 at *HLA-DPA1*, was a missense mutation at amino acid position 59 where Glu is converted to Asp. *HLA-DPA1*

as an HLA class II gene plays a central role in the immune system by presenting peptides derived from extracellular proteins. The other male-specific longevity SNP rs16981095 regulated the expression of the gene *TPM4* in whole blood (Westra et al., 2013) that its related pathways are dilated cardiomyopathy and cardiac muscle contraction. We also found four female-specific loci for longevity: rs10490092 at *FLJ30838*; rs6967652 at *BRAF*; rs73329134 at *BMPER*; and rs2075650 at *TOMM40/APOE*. Among these female longevity SNPs, rs2075650 had been linked to longevity in multiple studies as well as in our sex-combined analysis ($p = 6.17 \times 10^{-10}$). However, we found that rs2075650 is associated with longevity with higher effect in females ($\beta = -0.360$; $p = 5.12 \times 10^{-9}$) than males ($\beta = -0.197$; $p = 0.008$). More specifically, the *TOMM40/APOE* locus (chr19:45.37 M-45.42 M) could be split into four LD blocks (Figure 3B; Table S8), among which the SNPs in block 2 (rs12972156, rs12972970, rs34342646, rs6857, rs71352238, rs2075650, rs34404554) and block 4 (rs769449, rs429358) showed sex differences ($p_{\text{difference}} < 0.05$), whereas the SNPs in the other two blocks did not.

We next investigated sex-specific pathways based on sex-specific loci. The best-fit p cutoffs of 0.0178 and 0.0004 according to PRSice software were used to select SNPs for pathway analyses in males and females, respectively. The selected longevity-associated SNPs were significantly enriched in 9 pathways for males (false discovery rate, FDR < 0.05 ; Table S9). These pathways were mainly enriched in DNA replication and mismatch repair-related pathways including mismatch repair, nucleotide excision repair, base excision repair, and pathways related to amino acid metabolism including ATP-binding cassette transporters (ABC transporters), beta-alanine metabolism pathway, arginine, and proline metabolism pathway. In females, 10 pathways were enriched and clustered into the cancer-related pathway, including glioma pathway, melanoma pathway, chronic myeloid leukemia pathway, JAK-STAT signaling pathway, B-cell receptor signaling, and pathways related to the metabolism of terpenoids such as terpenoid backbone biosynthesis, porphyrin and chlorophyll metabolism, valine leucine and isoleucine degradation, and glyoxylate and dicarboxylate metabolism (Table S9).

3.4 | Replication of previously identified loci for human longevity

For validation purposes, we investigated previously reported 1,881 SNPs in multiple GWAS longevity studies (SNPs resources listed in Table S2). After quality control, 1,305 of the 1,881 SNPs were available for investigation. Nine SNPs in three genes were well replicated in this study and showed statistical significance after multiple testing corrections (Table S10; $p < 3.8 \times 10^{-5} = 0.05/1,305$). These three replicated genes include *TOMM40/APOE* (rs2075650, rs6857, rs769449, rs429358, rs405509, rs4420638, and rs7412), *FGD3* (rs4573339), and *AKT1* (rs3803304). Additionally, 106 SNPs were replicated with nominal significance ($p < 0.05$). For example, three previously reported SNPs, rs16972414 (*PIK3C3*, $p = 0.015$), rs6721003 (*SCN7A*, $p = 0.0037$), and rs2149954 (*EBF1*, $p = 0.047$),



were replicated (Table S11). However, our previously reported SNPs in gene *IL6* and *ANKRD20A9P* associations (Zeng et al., 2016) did not pass quality control. Therefore, they cannot be replicated.

Notably, we replicated 11 out of the 22 previously reported sex-specific SNPs but not the rare SNPs (Table S12). These 11 associations were nominally significant in one sex ($p < 0.05$) but not significant in the other sex ($p > 0.05$), and the sex-differential $p < 0.05$. SNP rs4972778 at *KIAA1715* showed the most significant difference between sexes ($p_{\text{male}} = 0.49$; $p_{\text{female}} = 6.62 \times 10^{-6}$; $p_{\text{sex difference}} = 5.13 \times 10^{-4}$).

3.5 | Observational, PRS, and MR analyses identify the correlations of diseases with longevity

On the basis of detailed questionnaire information including sex, age, diseases, cognition, and lifestyle factors, we systematically analyzed the effects of these factors on longevity. First, the observational correlations of longevity with diseases were investigated (Table S13). Interestingly, we found that the most influential factors related to an increased probability of longevity were being female, exhibiting a lower education level, exhibiting a lower career status, not smoking, not drinking, and an absence of diseases such as hypertension, type 2 diabetes mellitus (T2D), cardiovascular disease (CVD), dyslipidemia, gastroenteric ulcer, arthritis or cholelithiasis. At the organ and tissue aging levels, we found that long-lived individuals are more likely to suffer from cataracts, glaucoma, and dementia diseases, and to exhibit lower activities of daily living (ADL) and Mini-Mental State Examination (MMSE) scores. In addition, we found that T2D and CVD were significantly inversely correlated with longevity in females.

Based on a total of 3,966 disease markers from the GWAS catalog genotyped in this study, we constructed the polygenic risk scores (PRS) of each participant for 87 disease traits, and we calculated their correlations with longevity. Seven nominally significant correlations were identified, but none of them passed the multiple testing adjusted threshold (Table S14; $p < 0.0006$). However, we observed that long-lived individuals tended to present lower polygenic risk scores for T2D ($p = 0.002$), stroke ($p = 0.007$), rheumatoid arthritis ($p = 0.007$), vitiligo ($p = 0.031$), breast cancer ($p = 0.046$), or CVD ($p = 0.049$) with nominal significance ($p < 0.05$).

To investigate whether longevity might have causal effects on these diseases and vice versa, we next applied the bidirectional Mendelian randomization (MR) method. We explored multiple settings for instrument strength ($p < 10^{-3}$, $p < 10^{-4}$, and $p < 10^{-5}$) with strict LD pruning ($r^2 < 0.2$). We first identified the causal effect of longevity on the disease. When using all three different thresholds as cutoff values for instrumental strength, we observed that longevity had a causal effect on reduced risks of CVD ($p_{\text{GCTA-GSMR}} = 1.16 \times 10^{-8}$) and T2D ($p_{\text{GCTA-GSMR}} = 1.91 \times 10^{-6}$). These two causal relationships were shown to be robust when the other three MR tests were performed ($p_{\text{inverse_variance_weighted}} = 2.35 \times 10^{-9}$, $p_{\text{weighted_median}} = 1.75 \times 10^{-4}$, and $p_{\text{MR-Egger}} = 1.23 \times 10^{-2}$ for decreased CVD risk; $p_{\text{inverse_variance_weighted}} = 1.76 \times 10^{-6}$, $p_{\text{weighted_median}} = 5.64 \times 10^{-3}$, and $p_{\text{MR-Egger}} = 3.33 \times 10^{-2}$ for decreased T2D risk), and there was no evidence of horizontal pleiotropy ($p_{\text{egger-intercept}} > 0.05$; Table S15). Moreover, longevity showed direct observational correlations with T2D ($p_{\text{observation}} = 2.54 \times 10^{-12}$) and CVD ($p_{\text{observation}} = 1.58 \times 10^{-7}$; Table 3). Both observational correlations and bidirectional MR confirmed the causal effects of longevity on the lower risks of CVD and T2D. While defining $p < 10^{-3}$ as cutoff, longevous

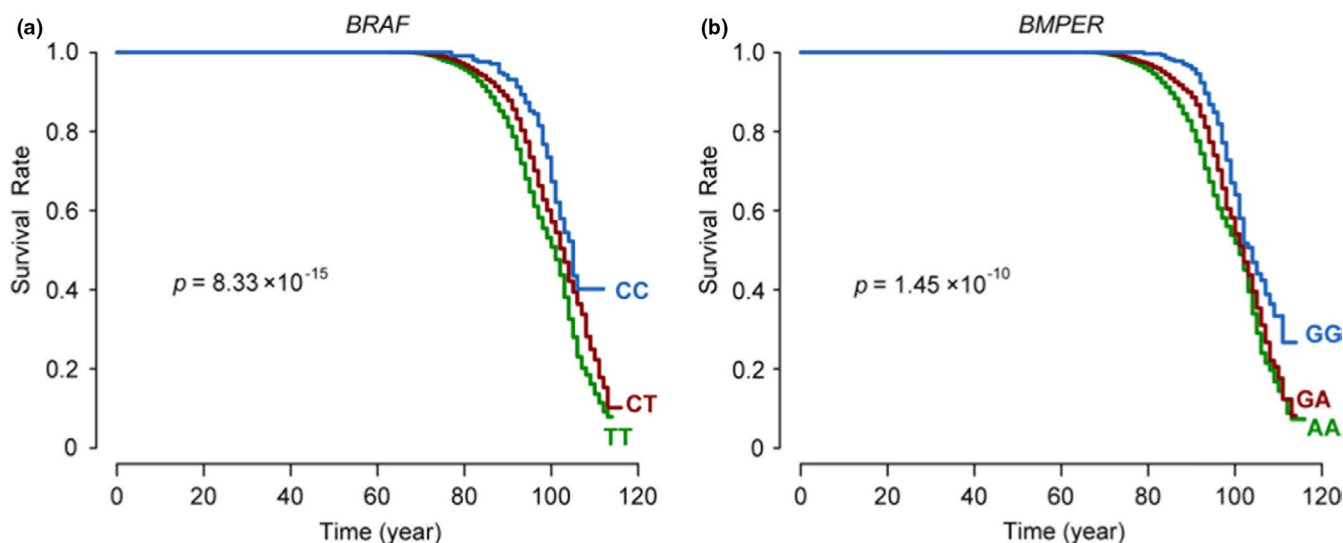


FIGURE 4 Survival curves for individuals with different genotypes. 12,664 individuals with live/ death status from dataset 1 were included for survival analysis. Approximately 24% (3,040/12,664) of the individuals died with age of death recorded. (a) Survival curves showing the life span of individuals according to *BRAF* gene polymorphism rs1267601. (b) Survival curves showing the life span of individuals according to *BMPER* gene polymorphism rs17169634.



individuals showed a causal effect on a decreased risk of arthritis ($p_{\text{GCTA-GSMR}} = 8.22 \times 10^{-5}$), stroke ($p_{\text{GCTA-GSMR}} = 2.57 \times 10^{-3}$), and hypertension ($p_{\text{GCTA-GSMR}} = 3.29 \times 10^{-3}$), presenting no horizontal pleiotropy. Conversely, we also investigated the potential effect of disease on longevity, and we observed that T2D and dementia negatively affected longevity when taking $p < 10^{-3}$ as an instrumental cutoff, while stroke and cataract negatively affected longevity when taking $p < 10^{-4}$ as the instrumental cutoff (Table S15). Taken together, our series of analyses, including observational correlation, PRS, and bidirectional MR analysis, revealed that long-lived individuals tend to exhibit a lower genetic risk of T2D, CVD, and arthritis, and in turn that the absence or delayed onset of diseases such as T2D, dementia, and stroke lead ones to live longer and have higher odds to be longevity (Table 3; Table S15).

3.6 | Determinants of life span identified by survival analysis

During the follow-ups, approximately 24% (3,040/12,664) of the individuals died (with an age of death recorded). We performed survival analyses to identify genetic variants associated with life span using the Cox regression model (Table S16). The results showed that rs1267601 in *BRAF* was the variant most associated with life span, and carriers of the CC genotype had significantly higher survival rates at age 100+ compared with CT/TT carriers (28.6% for CC, 16.9% for CT, and 11.8% for TT; hazard ratio (HR) for survival 1.35; $p = 8.33 \times 10^{-15}$; Figure 4A). In addition, we found that the Alzheimer's disease-related SNP rs17169634 in *BMPER* was the longevity marker showing the greatest promoting effect, with carriers

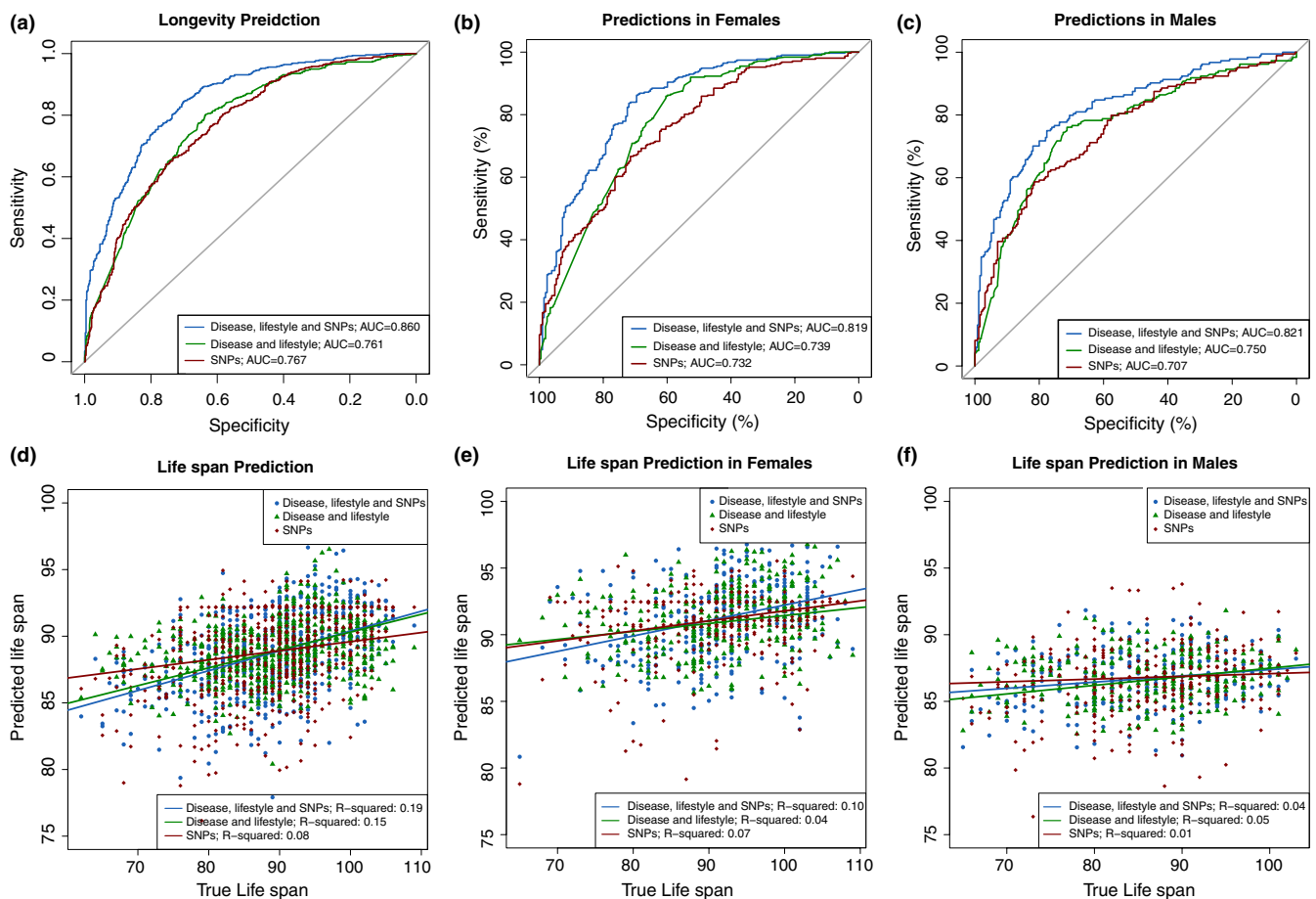


FIGURE 5 Predictive efficiency for longevity and life span. (a) Receiver operating characteristic curves (ROC) for predictor variables of longevity (binary phenotype). The predictive effectiveness was evaluated by the area under the ROC curve (AUC). The predictive effectiveness of SNPs (AUC = 0.767) was slightly higher compared with disease and lifestyle (AUC = 0.761). The composite SNPs, disease, and lifestyle achieved the best predictive power with AUC = 0.860. ROC for predictor variables of longevity in males (b) and in females (c), respectively. The predictive efficiencies of composite SNPs, diseases, and lifestyles for longevity were presented in both male (AUC = 0.821) and female (AUC = 0.819) groups. The predictive efficiency of SNPs for longevity is 0.732 in females and 0.707 in males. (d) The correlations between true life span and predicted life span showing the predicted effectiveness for life span. SNPs could only explain 2% of the variance for life span, while disease and lifestyle could explain much more variance (15.3%). The composite SNPs, disease, and lifestyle could explain 19.8% of the variance. The predicted effectiveness for life span in males (e) and in females (f), respectively. The predictive effectiveness for life span is better in females than in males. Especially, in males, the SNP set fails to give significant prediction for life span ($p = 0.10$), whereas the selected SNPs could explain 7% of the variance in females ($p = 1.25 \times 10^{-6}$).



of the minor allele G exhibiting a substantially longer life span than noncarrier (28.3% for GG, 16.8% for GA, and 11.3% for AA; survival HR = 1.25; $p = 1.45 \times 10^{-10}$; Figure 4B). Except for the *BRAF* and *BMPER* loci, we did not observe any other loci that reached genome-wide significance. The *APOE/TOMM40* locus was correlated with life span at nominal significance ($p = 0.0013$).

3.7 | Predictions of longevity and life span

One of the ultimate objectives for identifying factors contributing to longevity is to predict longevity and life span. Based on all the factors we identified from both previous studies and our own association studies, we constructed a predictive model for longevity (age ≥ 90 vs. age < 75) and life span through Lasso regression (Tibshirani, 1996). The prediction was independent of the association study; all the SNPs that we designed on the customized SNP chip ($n = 23,800$ after quality control) as well as 19 disease phenotypes and five lifestyles entered prediction model construction. We constructed three models using (1) all SNPs; (2) self-reported diseases and lifestyles collected in questionnaires; and (3) the combination of all features in models 1 and 2. After removing features and individuals with a high missing rate, the total sample size for the prediction study was 7807. Data were separated into training and testing sets (see Section 2). Models 1 and 2 achieved acceptable discriminations (model 1: AUC = 0.767; model 2: AUC = 0.761). Interestingly, model 1 (using all SNPs) and model 2 (using disease status and lifestyles) yielded similar prediction efficiency. Additionally, the third model, with all the variables, exhibited quite good predictive performances, and the final AUC reached 0.86 (Figure 5A). We further investigated the significance of the SNPs selected by Lasso regression for the prediction in our genetic association study. We found that those SNPs that effectively contributed to the prediction exhibited significantly lower p -value enrichment (Figure S6).

For the prediction of life span, we used 3,023 individuals who had an exact age of death and detailed phenotypic records in our datasets. All three predictive models yielded good performance (Figure 5; model 1: $p = 1.9 \times 10^{-5}$; models 2 and 3: $p < 2.2 \times 10^{-16}$). Model 1 could only explain 8% of the variances in the life span, while model 2 could explain 15.3%. Furthermore, model 3, including all features, could explain 19.8% of the variances.

The genetic architectures of longevity and disease progression are quite different between sexes. Next, we applied our predictive models to the male and female groups separately. The predictive efficiency for longevity (binary phenotype) was good in both the male (AUC = 0.821) and female (AUC = 0.819) groups (Figure 5b,c). However, the predictive efficiency for life span in females was superior to that of males (Figure 5e,f). In particular, in males, the SNP set failed to give a significant prediction for life span ($p = 0.10$), while in females, the selected SNPs could explain 7% of the variance ($p = 1.25 \times 10^{-6}$), indicating that the sex groups indeed exhibited different genetic architectures for aging. The selected features and their coefficients for all prediction models are listed in Tables S17 and S18.

4 | DISCUSSION

In this study, we present several findings regarding the genetic contributions to longevity and their gender differences based on 15,651 individuals from the cohort of the Chinese Longitudinal Healthy Longevity Survey. We designed an informative SNP chip for studying the genetics of longevity. Longevity case-control analysis ($n = 11,045$) and survival analysis ($n = 12,664$) were performed in different subsets of the cohort. In addition to previously published longevity-related studies, we included SNPs for relevant diseases, such as CVD and T2D, intending to obtain results that were comparable to those of previous studies. The main findings and several highlights of this work are described below.

First, we identified two novel loci (*BMPER* and *TMEM43/XPC*) and replicated three loci (*TOMM40/APOE*, *FGD3*, and *AKT1*) associated with longevity in Chinese populations. Interestingly, these five longevity-associated loci have been linked to diseases, especially age-related diseases. For example, *BMPER* has been associated with aging and its related diseases, such as Alzheimer's disease (Nelson et al., 2014), and it is also involved in the regulation of the proinflammatory phenotype of the endothelium (Helbing et al., 2011), functioning primarily in the vascular (Lockyer et al., 2017) and respiratory systems (Helbing et al., 2013). *XPC* is involved in DNA damage repair and is associated with disease characterized by an extreme sensitivity to ultraviolet rays from sunlight, such as xeroderma pigmentosum, complementation group c and xeroderma pigmentosum, variant type, and the deletion of *XPC* leads to lung tumors in mice (Hollander et al., 2005). The *TOMM40/APOE* locus has been reported to be associated with longevity in multiple studies among diverse populations and the locus contributes to Alzheimer's disease (Seshadri et al., 2010), age-related macular degeneration (Cipriani et al., 2012), cardiovascular disease (Middelberg et al., 2011), cognitive decline (Davies et al., 2014), immunity (Reiner et al., 2008), and lipid metabolism/dyslipidemia (Aulchenko et al., 2009). *FGD3*, a putative regulator of cell morphology and motility, was associated with longevity in the NECS study, and its expression plays a prognostic role in breast cancer (Renda et al., 2019). *AKT1* is relevant to longevity (Deelen et al., 2013; Nojima et al., 2013), and the dysregulation of AKT signaling leads to diseases for which there are major unmet medical needs, such as cancer, diabetes, and cardiovascular and neurological diseases (Hers et al., 2011). The two novel signals were also linked to multiple age-related phenotypes not only in this cohort but also in GWASs from the Japan BioBank, whose cohort is ethnically closer to the Chinese population. The *BMPER* locus was associated with arthritis, prostate cancer, and BMI. The *TMEM43* locus was associated with Parkinson's disease, AST, and UA. These findings consistently revealed the genetic overlap between exceptional longevity and age-related diseases and traits (Fortney et al., 2015). It is noted that five SNPs including rs10757274, rs4977574, rs2891168, rs10965235, and rs944797 located in well-known *CDKN2B* locus, which is associated with CVD, were also associated with longevity with nominal significance in our study.



While comparing our current results with other GWAS meta-analyses and our previous results, inconsistencies were found. SNP rs17169634 in *BMPER* showed different directions of effects in multiple studies. One possible reason could be that the true causal variants are hidden in this area, but due to different linkage disequilibrium (LD) structures among European and Chinese populations, the alleles tagging true causal variants may be different. Future fine-mapping with denser markers or genome sequencing will be required to illuminate the hidden information. SNP rs17169634 was not significantly associated with longevity in our dataset 2; therefore, the direction of effects in dataset 2 could not be determined that the confidence interval of effect size included zero. The significant signal was driven by dataset 1, where we also tested its association with complex diseases. The G allele for rs17169634 in *BMPER* has reduced effects on the risk of arthritis ($p = 3.76 \times 10^{-6}$) and prostate cancer ($p = 6.32 \times 10^{-3}$). Taken together, the results in our data showed that the G allele for rs17169634 in *BMPER* increased the probability of being longevity in our logistic regression and has increased effects for life expectancy in survival analysis and reduced the risks of age-related diseases. Notably, the causal effects from an increased chance of longevity to reduced risk of arthritis were also identified in our MR analysis. These directions of effects are as expected that long-lived individuals show a delay in overall morbidity through having beneficial effects for diseases (Andersen et al., 2012). As for our previously reported SNPs in *IL6* and *ANKRD20A9P*, they did not pass the quality control in our current analysis. Therefore, they cannot be replicated. We further checked the frequency of reported SNPs on these two genes. The minor allele of rs2069837 in *IL6* has a lower frequency (0.075) than it is in the dbSNP Asian population (0.179), and this allele has reduced effects for longevity. Since the proportion of centenarians is much higher in our previous study (48%; Zeng et al., 2016) than in other GWAS studies for longevity, the underrepresentation of this allele in our dataset is plausible. The inconsistency of the results could be caused by the differences in proportions of centenarians between our two datasets and also among different ethnic populations. As for SNP rs2440012 in *ANKRD20A9P*, the minor allele G was overrepresented in our previous study (0.076) compared with Asians in dbSNP (G = 0). It has been filtered out in Deelen's meta-analysis may be due to multi-allelic problems in the European population (C = 0.90, A = 0.0015, G = 0.098). Additional independent datasets are needed for a detailed look into these loci.

One existing major problem of longevity genetic studies is that the findings from different studies are difficult to replicate. The reasons could be ethnic differences in genetic background and variation of phenotype definitions. Furthermore, when considering the complex relationships between age-related diseases and longevity (Andersen et al., 2012; Ukraintseva et al., 2016), the health management systems and culture could also introduce distinctions among populations. For instance, if the age-related diseases could be managed well, the patients could still survive longer. Therefore, the longevous individuals are mixed with people who carry true protective alleles for the disease and individuals who accepted excellent health

care. While the healthcare systems are quite different among countries, especially for older generations, the frequency of longevity-related alleles could be different in the elderly. Hence, further studies should also take health care and lifestyle into account when classifying cases and controls for comparisons.

Second, we found some sex-specific loci related to longevity. Numerous studies have reported remarkable sex differences in longevity and life span (Austad & Fischer, 2016; Candore et al., 2006; Ostan et al., 2016; Yuan et al., 2020); however, very few studies have reported the sex-differential effect of genetics for longevity. *TOMM40/APOE* is well-characterized longevity locus that could be split into 4 LD blocks. We found that two of these 4 LD blocks were associated with longevity in females but not in males ($p_{\text{difference}} < 0.05$), in line with our previous study (Zeng et al., 2018). This may indicate sex-specific genetic associations of longevity may be caused by differences during meiosis between males and females. The distinction of recombination rates between sex groups has been reported in both human and animals (Li & Merila, 2010; Tapper et al., 2005). Since the recombination was closely interacted with natural selection (Schumer et al., 2018), differences in recombination are plausible to lead to sex or population stratification and thereby causing a small group of people having enriched evolutionary benefit alleles. Therefore, it is necessary to use strand-specific, long-segment sequencing technologies or family studies to detailed look into the LD structure for longevity people in future studies.

Interestingly, the predictive effectiveness of SNPs for longevity is slightly better in females (AUC = 0.732) than in males (AUC = 0.707). For life span predictions, SNPs could explain 7% of the variance for life span in females ($p = 1.25 \times 10^{-6}$) but failed to provide a significant prediction for life span in males ($p = 0.10$). All these results are consistent with our previous finding (Zeng et al., 2018) that the genetic association with longevity is stronger in females than in males. Notably, we found that some diseases also presented sex-differential patterns associated with longevity. For example, T2D and CVD were more significantly correlated with longevity in females. Previous studies have reported sex differences between cardiovascular diseases and aging, in which it is assumed that genetic traits and sex hormones play the key roles (Rodgers et al., 2019).

Our PRS and MR analyses revealed negative correlations between longevity and multiple diseases, including CVD, T2D, and arthritis. The results were generally in consistent with those in a meta-analysis of the European population (Joshi et al., 2017). However, other studies indicated different conclusions. One publication based on the Leiden Longevity Study (LLS) suggested that disease risk alleles do not compromise human longevity (Beekman et al., 2010). The authors only considered 30 disease risk SNPs, while our analyses included more carefully selected SNPs for age-related diseases (Erikson et al., 2016), and the obtained polygenic risk scores reflected an overall significant decrease in genetic disease risk in exceptionally long-lived individuals. Taken together, these findings suggested that some disease risk SNP alleles might increase the chance of longevity (McDaid et al., 2017), but there are more effective disease risk SNP alleles associated with earlier



mortality (Erikson et al., 2016; Joshi et al., 2017). The benefits of utilizing polygenic risk scores are that it summed the effects of multiple alleles instead of looking at the count of each risk allele. Therefore, when considering the additive effects aggregating all risk alleles, the genetic risks of multiple diseases were found to be reduced in long-lived populations.

There is growing interest in predicting the risks for diseases and complex traits using polygenic risk scores (Khera et al., 2018). Previous studies have predicted longevity and life span based mainly on animal models (Huang et al., 2004; Shen et al., 2014; Swindell et al., 2008) or the use of single biomarker (Ho et al., 2019; Whittemore et al., 2019). One recent study using the UK Biobank dataset applied polygenic risk score for life span prediction with good performance but was limited by utilizing summary statistics for life span based on parental data (Timmers et al., 2019). Our newly developed prediction models including genetic markers performed nicely in the classification of longevity (AUC = 0.767) but were not very effective in the prediction of life span (explaining only 15.3% of the variance). When disease status and lifestyle information were added, the longevity- and life span-prediction models produced better predictions for both longevity (AUC = 0.86) and life span (explaining 19.8% of the variances). Additionally, for the classification of longevity, the performance of SNPs is similar to that of diseases and lifestyles, indicating that genetics and phenotypes may have independent components that influence aging.

A limitation of the present study is the candidate-gene approach, which might preclude the discovery of new possible causative genes or biological pathways. However, our selection of candidates was primarily based on our previous genome-wide association studies conducted in 4477 Chinese individuals from the Chinese Longitudinal Healthy Longevity Survey (CLHLS). The SNPs with *p*-values smaller than 0.015 in our previous GWAS were all selected for inclusions on our customized chip. We collected additional candidate SNPs from existing studies, including studies not only on longevity genetics but also on other age-related complex diseases and traits. Moreover, by leveraging on imputation technology, the candidate SNP sets were further expanded. By incorporating all these SNPs together, we performed multi-candidate genes association analyses, which is suboptimal for genome-wide associations but still very informative. Secondly, in order to identify sex-specific genetic markers associated with longevity, we stratified our sample into male and female groups. The benefit of stratifying the sample is an increased chance to find those sex-specific SNPs tagging different causal variants in different sex groups. However, this strategy also has the drawback that the reduced sample size for each analysis group caused decreased power. We only replicated half of the previous identified sex-specific loci, and more replication studies are required in the future. Thirdly, we noted that the predicted life span was generally shorter than the true life span, indicating undefined missing confounders contributing to the life span (genetic and other confounding factors or their interactions). Future genetic studies of longevity based on affordable exome and whole-genome sequencing might be helpful to further identify a larger number of longevity-associated

genetic variants by applying the analysis of rare genetic and copy number variants. Together, these findings provide a benchmark for the development of longevity- and life span-predictive models. Further studies are warranted to improve the models through the identification of an additional panel of predictive variables and the development of innovative computational approaches.

In summary, our results not only identified novel longevity genes but also depicted the landscape of genetic contributors to longevity and life span through a complex of sex-differential and disease-related interactive circuits, which could be more precisely predicted in the near future.

ACKNOWLEDGMENTS

We thank for the support from the National Key Research and Development Program of China (2018YFC2000400, 2020YFC2002902 and 2018YFA0507800), the National Natural Science Foundation of China (No. 72061137004, No. 81903392, and No. 81941021), the Guangdong Provincial Key Laboratory of Genome Read and Write (2017B030301011), the Science, Technology and Innovation Commission of Shenzhen Municipality (No. JCYJ20170412153100794 and No. JCYJ20180507183615145), and the U.S. National Institute of Aging of National Institute of Health (P01AG031719).

CONFLICT OF INTEREST

We declare no competing interests.

AUTHOR CONTRIBUTIONS

YZ, CN, JM, and XX designed and supervised the study. Xiaomin Liu, ZS, and YL analyzed the data. Xiaomin Liu, ZS, YL, YY, and JM interpreted the data. MF, CB, PA, HC, ZC, BT, JS, XG, MZ, PC, TZ, HJ, Xiao Liu, YH, HY, JW, and FW were involved in data acquisition, collection of phenotype data, genotyping of the individuals, and data management. Xiaomin Liu, ZS, YL, YY, and JM wrote the manuscript. All authors read and approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study have been deposited and controlled in the CNSA (<https://db.cngb.org/cnsa/>) of CNGBdb with accession code CNP0000792. The data are currently not publicly available.

ORCID

Xiaomin Liu  <https://orcid.org/0000-0003-2155-3627>

Junxia Min  <https://orcid.org/0000-0001-8099-6327>

REFERENCES

- Altmann-Schneider, I., van der Grond, J., Slagboom, P. E., Westendorp, R. G., Maier, A. B., van Buchem, M. A., & de Craen, A. J. (2013). Lower susceptibility to cerebral small vessel disease in human familial longevity: The Leiden Longevity Study. *Stroke*, 44(1), 9–14. <https://doi.org/10.1161/STROKEAHA.112.671438>
- Andersen, S. L., Sebastiani, P., Dworkis, D. A., Feldman, L., & Perls, T. T. (2012). Health span approximates life span among many



- supercentenarians: Compression of morbidity at the approximate limit of life span. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 67(4), 395–405. <https://doi.org/10.1093/geron/a/glr223>
- Aulchenko, Y. S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I. M., Pramstaller, P. P., Penninx, B. W. J. H., Janssens, A. C. J. W., Wilson, J. F., Spector, T., Martin, N. G., Pedersen, N. L., Kyvik, K. O., Kaprio, J., Hofman, A., Freimer, N. B., Jarvelin, M.-R., Gyllenstein, U., Campbell, H., ... and for the ENGAGE Consortium (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nature Genetics*, 41(1), 47–55. <https://doi.org/10.1038/ng.269>
- Austad, S. N., & Fischer, K. E. (2016). Sex differences in lifespan. *Cell Metabolism*, 23(6), 1022–1033. <https://doi.org/10.1016/j.cmet.2016.05.019>
- Bae, H. T., Sebastiani, P., Sun, J. X., Andersen, S. L., Daw, E. W., Terracciano, A., Ferrucci, L., & Perls, T. T. (2013). Genome-wide association study of personality traits in the long life family study. *Frontiers in Genetics*, 4, 65. <https://doi.org/10.3389/fgene.2013.00065>
- Beekman, M., Nederstigt, C., Suchiman, H. E. D., Kremer, D., van der Breggen, R., Lakenberg, N., Alemayehu, W. G., de Craen, A. J. M., Westendorp, R. G. J., Boomsma, D. I., de Geus, E. J. C., Houwing-Duistermaat, J. J., Heijmans, B. T., & Slagboom, P. E. (2010). Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(42), 18046–18049. <https://doi.org/10.1073/pnas.1003540107>
- Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, 103(3), 338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Candore, G., Balistreri, C. R., Listi, F., Grimaldi, M. P., Vasto, S., Colonna-romano, G., Franceschi, C., Lio, D., Caselli, G., & Caruso, C. (2006). Immunogenetics, gender, and longevity. *Annals of the New York Academy of Sciences*, 1089, 516–537. <https://doi.org/10.1196/annals.1386.051>
- Cipriani, V., Leung, H.-T., Plagnol, V., Bunce, C., Khan, J. C., Shahid, H., Moore, A. T., Harding, S. P., Bishop, P. N., Hayward, C., Campbell, S., Armbricht, A. M., Dhillon, B., Deary, I. J., Campbell, H., Dunlop, M., Dominiczak, A. F., Mann, S. S., Jenkins, S. A., ... Yates, J. R. W. (2012). Genome-wide association study of age-related macular degeneration identifies associated variants in the TNXB-FKBPL-NOTCH4 region of chromosome 6p21.3. *Human Molecular Genetics*, 21(18), 4138–4150. <https://doi.org/10.1093/hmg/dds225>
- Davies, G., Harris, S. E., Reynolds, C. A., Payton, A., Knight, H. M., Liewald, D. C., Lopez, L. M., Luciano, M., Gow, A. J., Corley, J., Henderson, R., Murray, C., Pattie, A., Fox, H. C., Redmond, P., Lutz, M. W., Chiba-Falek, O., Linnertz, C., Saith, S., ... Deary, I. J. (2014). A genome-wide association study implicates the APOE locus in non-pathological cognitive ageing. *Molecular Psychiatry*, 19(1), 76–87. <https://doi.org/10.1038/mp.2012.159>
- de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Computational Biology*, 11(4), e1004219. <https://doi.org/10.1371/journal.pcbi.1004219>
- Deelen, J., Beekman, M., Uh, H.-W., Broer, L., Ayers, K. L., Tan, Q., Kamatani, Y., Bennet, A. M., Tamm, R., Trompet, S., Guðbjartsson, D. F., Flachsbarth, F., Rose, G., Viktorin, A., Fischer, K., Nygaard, M., Cordell, H. J., Crocco, P., van den Akker, E. B., ... Slagboom, P. E. (2014). Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Human Molecular Genetics*, 23(16), 4420–4432. <https://doi.org/10.1093/hmg/ddu139>
- Deelen, J., Evans, D. S., Arking, D. E., Tesi, N., Nygaard, M., Liu, X., Wojczynski, M. K., Biggs, M. L., van der Spek, A., Atzmon, G., Ware, E. B., Sarnowski, C., Smith, A. V., Seppälä, I., Cordell, H. J., Dose, J., Amin, N., Arnold, A. M., Ayers, K. L., ... Murabito, J. M. (2019). A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nature Communications*, 10(1), 3669. <https://doi.org/10.1038/s41467-019-11558-2>
- Deelen, J., Uh, H.-W., Monajemi, R., van Heemst, D., Thijssen, P. E., Böhringer, S., van den Akker, E. B., de Craen, A. J. M., Rivadeneira, F., Uitterlinden, A. G., Westendorp, R. G. J., Goeman, J. J., Slagboom, P. E., Houwing-Duistermaat, J. J., & Beekman, M. (2013). Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. *AGE*, 35(1), 235–249. <https://doi.org/10.1007/s11357-011-9340-3>
- Delaneau, O., Marchini, J., & Zagury, J. F. (2011). A linear complexity phasing method for thousands of genomes. *Nature Methods*, 9(2), 179–181. <https://doi.org/10.1038/nmeth.1785>
- Erikson, G. A., Bodian, D. L., Rueda, M., Molparia, B., Scott, E. R., Scott-Van Zeeland, A. A., Topol, S. E., Wineinger, N. E., Niederhuber, J. E., Topol, E. J., & Torkamani, A. (2016). Whole-genome sequencing of a healthy aging cohort. *Cell*, 165(4), 1002–1011. <https://doi.org/10.1016/j.cell.2016.03.022>
- Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic risk score software. *Bioinformatics*, 31(9), 1466–1468. <https://doi.org/10.1093/bioinformatics/btu848>
- Fortney, K., Dobriban, E., Garagnani, P., Pirazzini, C., Monti, D., Mari, D., Atzmon, G., Barzilay, N., Franceschi, C., Owen, A. B., & Kim, S. K. (2015). Genome-wide scan informed by age-related disease identifies loci for exceptional human longevity. *PLoS Genetics*, 11(12), e1005728. <https://doi.org/10.1371/journal.pgen.1005728>
- Greene, V. W. (2001). Personal hygiene and life expectancy improvements since 1850: Historic and epidemiologic associations. *American Journal of Infection Control*, 29(4), 203–206. <https://doi.org/10.1067/mic.2001.115686>
- Hammond, E. C., Garfinkel, L., & Seidman, H. (1971). Longevity of parents and grandparents in relation to coronary heart disease and associated variables. *Circulation*, 43(1), 31–44. <https://doi.org/10.1161/01.cir.43.1.31>
- Helbing, T., Herold, E.-M., Hornstein, A., Wintrich, S., Heinke, J., Grundmann, S., Patterson, C., Bode, C., & Moser, M. (2013). Inhibition of BMP activity protects epithelial barrier function in lung injury. *The Journal of Pathology*, 231(1), 105–116. <https://doi.org/10.1002/path.4215>
- Helbing, T., Rothweiler, R., Ketterer, E., Goetz, L., Heinke, J., Grundmann, S., Duerschmied, D., Patterson, C., Bode, C., & Moser, M. (2011). BMP activity controlled by BMPER regulates the proinflammatory phenotype of endothelium. *Blood*, 118(18), 5040–5049. <https://doi.org/10.1182/blood-2011-03-339762>
- Hers, I., Vincent, E. E., & Tavaré, J. M. (2011). Akt signalling in health and disease. *Cellular Signalling*, 23(10), 1515–1527. <https://doi.org/10.1016/j.cellsig.2011.05.004>
- Herskind, A. M., McGue, M., Holm, N. V., Sorensen, T. I., Harvald, B., & Vaupel, J. W. (1996). The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics*, 97(3), 319–323. <https://doi.org/10.1007/bf02185763>
- Hjelmborg, J. V. B., Iachine, I., Skytthe, A., Vaupel, J. W., McGue, M., Koskenvuo, M., Kaprio, J., Pedersen, N. L., & Christensen, K. (2006). Genetic influence on human lifespan and longevity. *Human Genetics*, 119(3), 312–321. <https://doi.org/10.1007/s00439-006-0144-y>
- Ho, A., Sinick, J., Esko, T., Fischer, K., Menni, C., Zierer, J., Matey-Hernandez, M., Fortney, K., & Morgen, E. K. (2019). Circulating glucuronic acid predicts healthspan and longevity in humans and mice. *Aging*, 11(18), 7694–7706. <https://doi.org/10.18632/aging.102281>
- Hollander, M. C., Philburn, R. T., Patterson, A. D., Velasco-Miguel, S., Friedberg, E. C., Linnoila, R. I., & Fornace, Jr., A. J. (2005). Deletion of XPC leads to lung tumors in mice and is associated with early events in human lung carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37), 13200–13205. <https://doi.org/10.1073/pnas.0503133102>



- Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, 5(6), e1000529. <https://doi.org/10.1371/journal.pgen.1000529>
- Huang, C., Xiong, C., & Kornfeld, K. (2004). Measurements of age-related changes of physiological processes that predict lifespan of *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(21), 8084–8089. <https://doi.org/10.1073/pnas.0400848101>
- Joshi, P. K., Pirastu, N., Kentistou, K. A., Fischer, K., Hofer, E., Schraut, K. E., Clark, D. W., Nutile, T., Barnes, C. L. K., Timmers, P. R. H. J., Shen, X., Gandin, I., McDaid, A. F., Hansen, T. F., Gordon, S. D., Giulianini, F., Boutin, T. S., Abdellaoui, A., Zhao, W., ... Wilson, J. F. (2017). Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nature Communications*, 8(1), 910. <https://doi.org/10.1038/s41467-017-00934-5>
- Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, 50(9), 1219–1224. <https://doi.org/10.1038/s41588-018-0183-z>
- Kitamura, H., Kobayashi, Y., Kimura, A., & Numano, F. (1998). Association of clinical manifestations with HLA-B alleles in Takayasu arteritis. *International Journal of Cardiology*, 66(Suppl. 1), S121–126.
- Li, M. H., & Merila, J. (2010). Sex-specific population structure, natural selection, and linkage disequilibrium in a wild bird population as revealed by genome-wide microsatellite analyses. *BMC Evolutionary Biology*, 10, 66. <https://doi.org/10.1186/1471-2148-10-66>
- Lockyer, P., Mao, H., Fan, Q., Li, L., Yu-Lee, L.-Y., Eissa, N. T., Patterson, C., Xie, L., & Pi, X. (2017). LRP1-dependent BMPER signaling regulates lipopolysaccharide-induced vascular inflammation. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 37(8), 1524–1535. <https://doi.org/10.1161/ATVBAHA.117.309521>
- Lunetta, K. L., D'Agostino, R. B., Karasik, D., Benjamin, E. J., Guo, C.-Y., Govindaraju, R., Kiel, D. P., Kelly-Hayes, M., Massaro, J. M., Pencina, M. J., Seshadri, S., & Murabito, J. M. (2007). Genetic correlates of longevity and selected age-related phenotypes: A genome-wide association study in the Framingham Study. *BMC Medical Genetics*, 8(Suppl. 1), S13. <https://doi.org/10.1186/1471-2350-8-S1-S13>
- Mabuchi, T., Ota, T., Manabe, Y., Ikoma, N., Ozawa, A., Terui, T., & Oka, A. (2014). HLA-C*12:02 is a susceptibility factor in late-onset type of psoriasis in Japanese. *Journal of Dermatology*, 41(8), 697–704. <https://doi.org/10.1111/1346-8138.12569>
- MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z. M., Welter, D., Burdett, T., Hindorf, L., Flicek, P., Cunningham, F., & Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, 45(D1), D896–D901. <https://doi.org/10.1093/nar/gkw1133>
- McDaid, A. F., Joshi, P. K., Porcu, E., Komljenovic, A., Li, H., Sorrentino, V., Litovchenko, M., Bevers, R. P. J., Rüeger, S., Reymond, A., Bochud, M., Deplancke, B., Williams, R. W., Robinson-Rechavi, M., Paccaud, F., Rousson, V., Auwerx, J., Wilson, J. F., & Kutalik, Z. (2017). Bayesian association scan reveals loci associated with human lifespan and linked biomarkers. *Nature Communications*, 8, 15842. <https://doi.org/10.1038/ncomms15842>
- Middelberg, R. P. S., Ferreira, M. A. R., Henders, A. K., Heath, A. C., Madden, P. A. F., Montgomery, G. W., Martin, N. G., & Whitfield, J. B. (2011). Genetic variants in LPL, OASL and TOMM40/APOE-C1-C2-C4 genes are associated with multiple cardiovascular-related traits. *BMC Medical Genetics*, 12, 123. <https://doi.org/10.1186/1471-2350-12-123>
- Nelson, P. T., Estus, S., Abner, E. L., Parikh, I., Malik, M., Neltner, J. H., Ighodaro, E., Wang, W.-X., Wilfred, B. R., Wang, L.-S., Kukull, W. A., Nandakumar, K., Farman, M. L., Poon, W. W., Corrada, M. M., Kawas, C. H., Cribbs, D. H., Bennett, D. A., Schneider, J. A., ... Fardo, D. W. (2014). ABCC9 gene polymorphism is associated with hippocampal sclerosis of aging pathology. *Acta Neuropathologica*, 127(6), 825–843. <https://doi.org/10.1007/s00401-014-1282-2>
- Nojima, A., Yamashita, M., Yoshida, Y., Shimizu, I., Ichimiya, H., Kamimura, N., Kobayashi, Y., Ohta, S., Ishii, N., & Minamino, T. (2013). Haploinsufficiency of akt1 prolongs the lifespan of mice. *PLoS One*, 8(7), e69178. <https://doi.org/10.1371/journal.pone.0069178>
- Oeppen, J., & Vaupel, J. W. (2002). Demography. Broken limits to life expectancy. *Science*, 296(5570), 1029–1031. <https://doi.org/10.1126/science.1069675>
- Ostan, R., Monti, D., Guerresi, P., Bussolotto, M., Franceschi, C., & Baggio, G. (2016). Gender, aging and longevity in humans: An update of an intriguing/neglected scenario paving the way to a gender-specific medicine. *Clinical Science*, 130(19), 1711–1725. <https://doi.org/10.1042/CS20160004>
- Passarino, G., De Rango, F., & Montesanto, A. (2016). Human longevity: Genetics or lifestyle? It takes two to tango. *Immun Ageing*, 13, 12. <https://doi.org/10.1186/s12979-016-0066-z>
- Perls, T., Shea-Drinkwater, M., Bowen-Flynn, J., Ridge, S. B., Kang, S., Joyce, E., Daly, M., Brewster, S. J., Kunkel, L., & Puca, A. A. (2000). Exceptional familial clustering for extreme longevity in humans. *Journal of the American Geriatrics Society*, 48(11), 1483–1485.
- Reiner, A. P., Barber, M. J., Guan, Y., Ridker, P. M., Lange, L. A., Chasman, D. I., & Krauss, R. M. (2008). Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *American Journal of Human Genetics*, 82(5), 1193–1201. <https://doi.org/10.1016/j.ajhg.2008.03.017>
- Renda, I., Bianchi, S., Vezzosi, V., Nori, J., Vanzi, E., Tavella, K., & Susini, T. (2019). Expression of FGD3 gene as prognostic factor in young breast cancer patients. *Scientific Reports*, 9(1), 15204. <https://doi.org/10.1038/s41598-019-51766-w>
- Rodgers, J. L., Jones, J., Bolleddu, S. I., Vanthenapalli, S., Rodgers, L. E., Shah, K., Karia, K., & Panguluri, S. K. (2019). Cardiovascular risks associated with gender and aging. *Journal of Cardiovascular Development and Disease*, 6(2). <https://doi.org/10.3390/jcdd6020019>
- Rosa, M., Chignon, A., Li, Z., Boulanger, M.-C., Arsenault, B. J., Bossé, Y., Thériault, S., & Mathieu, P. (2019). A Mendelian randomization study of IL6 signaling in cardiovascular diseases, immune-related disorders and longevity. *NPJ Genomic Medicine*, 4, 23. <https://doi.org/10.1038/s41525-019-0097-4>
- Sakaue, S., Kanai, M., Karjalainen, J., Akiyama, M., Kurki, M., Matoba, N., Takahashi, A., Hirata, M., Kubo, M., Matsuda, K., Murakami, Y., Daly, M. J., Kamatani, Y., & Okada, Y. (2020). Trans-biobank analysis with 676,000 individuals elucidates the association of polygenic risk scores of complex traits with human lifespan. *Nature Medicine*, 26(4), 542–548. <https://doi.org/10.1038/s41591-020-0785-8>
- Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, 360(6389), 656–660. <https://doi.org/10.1126/science.aar3684>
- Sebastiani, P., Gurinovich, A., Bae, H., Andersen, S., Malovini, A., Atzmon, G., Villa, F., Kraja, A. T., Ben-Avraham, D., Barzilai, N., Puca, A., & Perls, T. T. (2017). Four genome-wide association studies identify new extreme longevity variants. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 72(11), 1453–1464. <https://doi.org/10.1093/gerona/glx027>
- Sebastiani, P., Solovieff, N., DeWan, A. T., Walsh, K. M., Puca, A., Hartley, S. W., Melista, E., Andersen, S., Dworkis, D. A., Wilk, J. B., Myers, R. H., Steinberg, M. H., Montano, M., Baldwin, C. T., Hoh, J., & Perls, T. T. (2012). Genetic signatures of exceptional longevity in humans. *PLoS One*, 7(1), e29848. <https://doi.org/10.1371/journal.pone.0029848>
- Seshadri, S., Fitzpatrick, A. L., Ikram, M. A., DeStefano, A. L., Gudnason, V., Boada, M., Bis, J. C., Smith, A. V., Carassquillo, M. M., Lambert,



- J. C., Harold, D., Schrijvers, E. M. C., Ramirez-Lorca, R., DeBette, S., Longstreth, Jr., W. T., Janssens, A. C. J. W., Shane Pankratz, V., Dartigues, J. F., Hollingworth, P. ... CHARGE Consortium; GERAD1 Consortium; EADI1 Consortium. (2010). Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*, 303(18), 1832–1840. <https://doi.org/10.1001/jama.2010.574>
- Shen, E.-Z., Song, C.-Q., Lin, Y., Zhang, W.-H., Su, P.-F., Liu, W.-Y., Zhang, P., Xu, J., Lin, N. A., Zhan, C., Wang, X., Shyr, Y. U., Cheng, H., & Dong, M.-Q. (2014). Mitoflash frequency in early adulthood predicts lifespan in *Caenorhabditis elegans*. *Nature*, 508(7494), 128–132. <https://doi.org/10.1038/nature13012>
- Skyttke, A., Pedersen, N. L., Kaprio, J., Stazi, M. A., Hjelmberg, J. V. B., Lachine, I., Vaupel, J. W., & Christensen, K. (2003). Longevity studies in GenomEUtwin. *Twin Research*, 6(5), 448–454. <https://doi.org/10.1375/136905203770326457>
- Swindell, W. R., Harper, J. M., & Miller, R. A. (2008). How long will my mouse live? Machine learning approaches for prediction of mouse life span. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 63(9), 895–906. <https://doi.org/10.1093/gerona/63.9.895>
- Tanaka, T., Dutta, A., Pilling, L. C., Xue, L., Lunetta, K. L., Murabito, J. M., Bandinelli, S., Wallace, R., Melzer, D., & Ferrucci, L. (2017). Genome-wide association study of parental life span. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 72(10), 1407–1410. <https://doi.org/10.1093/gerona/glw206>
- Tapper, W., Collins, A., Gibson, J., Maniatis, N., Ennis, S., & Morton, N. E. (2005). A map of the human genome in linkage disequilibrium units. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33), 11835–11839. <https://doi.org/10.1073/pnas.0505262102>
- Terry, D. F., Evans, J. C., Pencina, M. J., Murabito, J. M., Vasan, R. S., Wolf, P. A., & Benjamin, E. J. (2007). Characteristics of Framingham offspring participants with long-lived parents. *Archives of Internal Medicine*, 167(5), 438–444. <https://doi.org/10.1001/archinte.167.5.438>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
- Timmers, P. R. H. J., Mounier, N., Lall, K., Fischer, K., Ning, Z., Feng, X., Bretherick, A. D., Clark, D. W., Agbessi, M., Ahsan, H., Alves, I., Andiappan, A., Awadalla, P., Battle, A., Bonder, M. J., Boomsma, D., Christiansen, M., Claringbould, A., Deelen, P., ... Joshi, P. K. (2019). Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife*, 8, <https://doi.org/10.7554/eLife.39856>
- Triendl, R. (2003). Japan launches controversial Biobank project. *Nature Medicine*, 9(8), 982. <https://doi.org/10.1038/nm0803-982b>
- Ukrainitseva, S., Yashin, A., Arbeeve, K., Kulminski, A., Akushevich, I., Wu, D., Joshi, G., Land, K. C., & Stallard, E. (2016). Puzzling role of genetic risk factors in human longevity: "Risk alleles" as pro-longevity variants. *Biogerontology*, 17(1), 109–127. <https://doi.org/10.1007/s10522-015-9600-1>
- United Nations, D. o. E. a. S. A., Population Division. (2019). World Population Prospects 2019: Highlights. ST/ESA/SER.A/423.
- van den Berg, N., Rodríguez-Girondo, M., van Dijk, I. K., Mourits, R. J., Mandemakers, K., Janssens, A. A. P. O., Beekman, M., Smith, K. R., & Slagboom, P. E. (2019). Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nature Communications*, 10(1), 35. <https://doi.org/10.1038/s41467-018-07925-0>
- van der Lee, S. J., Conway, O. J., Jansen, I., Carrasquillo, M. M., Kleiendam, L., van den Akker, E., Hernández, I., van Eijk, K. R., Stringa, N., Chen, J. A., Zettergren, A., Andlauer, T. F. M., Diez-Fairen, M., Simon-Sanchez, J., Lleó, A., Zetterberg, H., Nygaard, M., Blauwendraat, C., Savage, J. E., ... Holstege, H. (2019). A nonsynonymous mutation in PFCG2 reduces the risk of Alzheimer's disease, dementia with Lewy bodies and frontotemporal dementia, and increases the likelihood of longevity. *Acta Neuropathologica*, 138(2), 237–250. <https://doi.org/10.1007/s00401-019-02026-8>
- Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, 8(1), 1826. <https://doi.org/10.1038/s41467-017-01261-5>
- Westra, H.-J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M. W., Fairfax, B. P., Schramm, K., Powell, J. E., Zernakova, A., Zernakova, D. V., Veldink, J. H., Van den Berg, L. H., Karjalainen, J., Withoff, S., Uitterlinden, A. G., Hofman, A., Rivadeneira, F., ... Franke, L. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics*, 45(10), 1238–1243. <https://doi.org/10.1038/ng.2756>
- Whittemore, K., Vera, E., Martinez-Nevedo, E., Sanpera, C., & Blasco, M. A. (2019). Telomere shortening rate predicts species life span. *Proceedings of the National Academy of Sciences of the United States of America*, 116(30), 15122–15127. <https://doi.org/10.1073/pnas.1902452116>
- Yashin, A. I., Arbeeve, K. G., Wu, D., Arbeeve, L. S., Bagley, O., Stallard, E., Kulminski, A. M., Akushevich, I., Fang, F., Wojczynski, M. K., Christensen, K., Newman, A. B., Boudreau, R. M., Province, M. A., Thielke, S., Perls, T. T., An, P., Elo, I., & Ukrainitseva, S. V. (2018). Genetics of human longevity from incomplete data: New findings from the long life family study. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 73(11), 1472–1481. <https://doi.org/10.1093/gerona/gly057>
- Yuan, R., Musters, C. J. M., Zhu, Y., Evans, T. R., Sun, Y., Chesler, E. J., & Bartke, A. (2020). Genetic differences and longevity-related phenotypes influence lifespan and lifespan variation in a sex-specific manner in mice. *Aging Cell*, e13263, <https://doi.org/10.1111/accel.13263>
- Zeng, Y. I., Nie, C., Min, J., Chen, H., Liu, X., Ye, R., Chen, Z., Bai, C., Xie, E., Yin, Z., Lv, Y., Lu, J., Li, J., Ni, T., Bolund, L., Land, K. C., Yashin, A., O'Rand, A. M., Sun, L., ... Vaupel, J. (2018). Sex differences in genetic associations with longevity. *JAMA Network Open*, 1(4), <https://doi.org/10.1001/jamanetworkopen.2018.1670>
- Zeng, Y. I., Nie, C., Min, J., Liu, X., Li, M., Chen, H., Xu, H., Wang, M., Ni, T., Li, Y., Yan, H., Zhang, J.-P., Song, C., Chi, L.-Q., Wang, H.-M., Dong, J., Zheng, G.-Y., Lin, L. I., Qian, F., ... Vaupel, J. W. (2016). Novel loci and pathways significantly associated with longevity. *Scientific Reports*, 6, 21243. <https://doi.org/10.1038/srep21243>
- Zhou, F., Cao, H., Zuo, X., Zhang, T., Zhang, X., Liu, X., Xu, R., Chen, G., Zhang, Y., Zheng, X., Jin, X., Gao, J., Mei, J., Sheng, Y., Li, Q., Liang, B. O., Shen, J., Shen, C., Jiang, H., ... Zhang, X. (2016). Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. *Nature Genetics*, 48(7), 740–746. <https://doi.org/10.1038/ng.3576>
- Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M. R., McGrath, J. J., Visscher, P. M., Wray, N. R., & Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications*, 9(1), 224. <https://doi.org/10.1038/s41467-017-02317-2>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Liu X, Song Z, Li Y, et al. Integrated genetic analyses revealed novel human longevity loci and reduced risks of multiple diseases in a cohort study of 15,651 Chinese individuals. *Aging Cell*. 2021;20:e13323. <https://doi.org/10.1111/accel.13323>