

Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research

Konrad-Zuse-Strasse 1 • D-18057 Rostock • Germany • Tel +49 (0) 3 81 20 81 - 0 • Fax +49 (0) 3 81 20 81 - 202 • www.demogr.mpg.de

MPIDR Working Paper WP 2017-004 | February 2017

**Sex differences in genetic associations
with longevity in Han Chinese: Sex-
stratified genome-wide association
study and polygenic risk score analysis**

Yi Zeng | zengyi@duke.edu

Huashuai Chen

Xiaomin Liu

Rui Ye

Enjun Xie

Zhihua Chen

Jiehua Lu

Jianxin Li

Yaohua Tian

Ting Ni

Lars Bolund

Kenneth C. Land

Anatoliy Yashin

Angela M O'Rand

Liang Sun

Ze Yang

Wei Tao

Anastasia Gurinovich

Claudio Franceschi

Jichun Xie

Jun Gu

Yong Hou

Xiao Liu

Xun Xu

Jean-Marie Robine

Joris Deelen

Paola Sebastiani

Eline Slagboom

Thomas Perls

Elizabeth Hauser

William Gottschalk

Qihua Tan

Kaare Christensen

Mike Lutz

Xiao-Li Tian

Huanming Yang

Junxia Min | junxiamin@zju.edu.cn

Chao Nie | niechao@genomics.cn

James Vaupel | jvv@demogr.mpg.de

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

Sex differences in genetic associations with longevity in Han Chinese:

Sex-stratified genome-wide association study and polygenic risk score analysis

Yi Zeng^{a,b,1,2}, Huashuai Chen^{a,c,1}, Xiaomin Liu^{d,1}, Rui Ye^{d,1}, Enjun Xie^e, Zhihua Chen^d, Jiehua Lu^f, Jianxin Li^f, Yaohua Tian^g, Ting Ni^h, Lars Bolund^{d,i}, Kenneth C. Land^j, Anatoliy Yashinⁱ, Angela M O'Rand^j, Liang Sun^k, Ze Yang^k, Wei Tao^l, Anastasia Gurinovich^m, Claudio Franceschiⁿ, Jichun Xie^o, Jun Gu^l, Yong Hou^d, Xiao Liu^d, Xun Xu^d, Jean-Marie Robine^p, Joris Deelen^q, Paola Sebastiani^m, Eline Slagboom^r, Thomas Perls^m, Elizabeth Hauser^s, William Gottschalk^t, Qihua Tan^u, Kaare Christensen^u, Mike Lutz^t, Xiao-Li Tian^v, Huanming Yang^{d,w}, Junxia Min^{e,2}, Chao Nie^{d,2}, James Vaupel^{x,2}

^a Center for the Study of Aging and Human Development, Medical School of Duke University, Durham, NC, USA, 27710. ^b Center for Healthy Aging and Development Studies, National School of Development, Peking University, Beijing, China, 100871. ^c Business School of Xiangtan University, Xiangtan, China, 411105. ^d BGI-Shenzhen, Shenzhen, China, 518083. ^e Institute of Translational Medicine, School of Medicine, Zhejiang University, Hangzhou, China, 311058. ^f Department of Sociology, Peking University, Beijing, China, 100871. ^g School of Public Health, Peking University, Beijing, China, 100871. ^h School of Life Sciences, Fudan University, Shanghai, China, 200433. ⁱ Department of Biomedicine, Aarhus University, Aarhus, Denmark, 8000. ^j Population Research Institute, Duke University, Durham, NC, USA, 27710. ^k National Institutes of Geriatrics, Beijing Hospital, Beijing, China, 100730. ^l School of Life Sciences, Peking University, Beijing, China, 100871. ^m Boston University, Boston, MA, 02215. ⁿ University of Bologna, Bologna, Italy, 40126. ^o Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA, 27710. ^p INSERM and EPHE, University of Montpellier, Montpellier, France, 34090. ^q Max Planck Institute for Biology of Ageing, Cologne, Germany, 50931. ^r Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, 2333. ^s Molecular Physiology Institute, Medical Center, Duke University, Durham, NC, USA, 27710. ^t Department of Neurology, Medical Center, Duke University, Durham, NC, USA, 27710. ^u University of Southern Denmark, Odense, Denmark, 6400. ^v Human Aging Research Institute and School of Life Science, Nanchang University, Jiangxi, China, 330031. ^w James D. Watson Institute of Genome Sciences, Hangzhou 310008, China, 310058. ^x Max Planck Institute for Demographic Research, Rostock, Germany, 50931.

¹ These authors contributed equally in analyzing data and interpreting results.

² These authors jointly supervised this work.

Address correspondences to: zengyi@duke.edu; junxiamin@zju.edu.cn; niechao@genomics.cn; JWV@demogr.mpg.de

Key Words: longevity, centenarians, sex-stratified GWAS, sex-specific pathways, polygenic risk scores, bio-demography

Abstract

Based on sex-stratified genome-wide association study (GWAS) of Han Chinese, 2,178 centenarians and 2,299 middle-aged controls, we identified 11 male- and 12 female-specific independent loci that are significantly associated with longevity ($P < 10^{-5}$), replicated in independent North and South regions in one sex, but are not significant ($P > 0.05$) in the other sex. We found that the association of rs60210535 at *LINC00871* with longevity replicated well between Chinese females ($P = 4.6 \times 10^{-5}$) and U.S. females ($P = 9.0 \times 10^{-5}$), but was not significant in both Chinese and U.S. males ($P > 0.05$). We discovered that 11 male-specific and 34 female-specific pathways are significantly associated with longevity ($P < 0.005$, false discovery rate (FDR) < 0.05). Male-specific pathways are enriched for inflammation and immunity genes, but female-specific pathways include tryptophan metabolic and PGC-1 α pathways that converge to mitochondrial biogenesis. Polygenic risk score (PRS) analyses demonstrated that 11/12 male/female top loci ($P < 10^{-5}$ in one sex, $P > 0.05$ in other sex), 44/58 male/female strong loci ($10^{-5} \leq P < 10^{-4}$ in one sex, $P > 0.4$ or $P > 0.35$ in other sex), and 191/311 male/female moderate loci ($10^{-4} \leq P < 10^{-3}$ in one sex, $P > 0.75$ or $P > 0.7$ in other sex) are jointly and highly associated with longevity exceeding a significance level $P < 10^{-8}$ in one sex, but not jointly associated with longevity in the other sex ($P > 0.05$). Our integrated PRS and novel sex-specific genetic relative benefit/loss ratio analyses indicate that females' genetic constitution favors longevity more than males'. Further interdisciplinary collaborative efforts are warranted, such as replications from other populations, international meta-analyses with much larger sample size, lab tests, and *in silico* functional validations.

Significance Statement: On average, women live significantly longer lives than men but they have lower physical performance and more adverse health outcomes at older ages compared to men: patterns that signify the *male-female health-survival paradox*

(1). Research on sex differences in health and mortality has proliferated, but has yet to achieve a good understanding of the effects of genetic variants on the sex gap in longevity and health. Based on sex-stratified genome-wide association analysis (GWAS) of Han Chinese including centenarians with a sample size 2.7 times as large as other published largest single GWAS on longevity involving centenarians (2), the present study aims to contribute a better understanding of sex differences in genetic associations with longevity.

\body

INTRODUCTION

Existing literature indicates that the associations of some specific genetic variants with health outcomes differ between men and women. For example, a significantly higher cumulative genetic risk for systemic lupus erythematosus was observed in men than in women (3). A recent study summarized 33 autosomal loci that have sex-specific effects on 22 traits including Crohn's disease, type 2 diabetes mellitus, plasma homocysteine, high density lipoprotein and thyroid stimulating hormone (4). It was found that the risks associated with the apolipoprotein E4 allele and the brain derived neurotrophic factor (*BDNF*) Met66 allele are greater in women compared to men (5,6).

It is widely recognized that centenarian genomes may harbor genetic variants associated with health and longevity and thus all of the GWAS on longevity use centenarians (and nonagenarians; a few used aged 85+) as cases and middle-aged or young-old adults as controls (7,8) (section S1 of Supplemental Materials (SM)). All previously published GWAS on longevity employed sexes-combined datasets adjusted for sex as a covariate (2,7,9,10). A few longevity GWAS studies conducted sex-specific analyses on the significant loci that were identified and replicated in the sexes-combined discovery and evaluation stages, but none of those studies found that replicated loci had significant sex differences in the association with longevity (2,7,9,10). These results are expected statistically, because the tested variable cannot

be significant and replicated in the sexes-combined datasets if it is significant in one sex but not significant in the other sex, given that the sub-sample size of either one of the two sexes is usually not small as to affect the overall results in sexes-combined dataset (11). In other words, the genetic variants associated with longevity identified in all previously published GWAS are sex-independent.

To identify sex-specific genetic variants associated with longevity, we conducted the first sex-stratified GWAS using Han Chinese independent datasets of 564 male and 1614 female centenarians and 773 male and 1526 female middle-aged controls from the Chinese Longitudinal Healthy Longevity Study (CLHLS) (SM Table 1). There were approximately five centenarians per million in China in the 1990s, compared with 50 per million in Western Europe in the same period (12). Plausibly, Han Chinese centenarians may be more likely to have longevity-associated genes than their centenarian counterparts in the Western world since they survived more brutal mortality regimes of the past when famine, wars, and starvation operated on birth cohorts of many millions. And unlike Western countries that received many international immigrants from other parts of the world resulting in relatively heterogeneous genetic compositions even within the same ethnic group, China has received very few international immigrants. Consequently, Han Chinese (consisting of about 93 percent of the total population in China) are relatively more homogenous in genetic composition compared to their Western counterparts (13). Thus, this sex-stratified GWAS with large samples of Han Chinese male and female centenarians is expected to be instrumental for understanding sex differences in genetic associations with longevity.

RESULTS

Analytic Overview

We use the CLHLS GWAS datasets which were published recently (10) (SM section S2). After standard GWAS quality-control filtering for subjects and single nucleotide

polymorphisms (SNP) as described in the reference (10), 5.6 million SNPs (0.82 million genotyped SNPs and 4.8 million imputed SNPs) in 564 male and 1614 female centenarians and 773 male and 1526 female middle-aged controls form the basis for our sex-stratified GWAS. Based on existing literature on Chinese genetic studies (13) and principal component analysis (10), we stratified the sex-specific samples into independent North and South regions as discovery and evaluation datasets (SM Table 1). The genomic inflation factors (λ) in the South, North and combined datasets were 1.022, 1.010 and 1.022, respectively, indicating that the effects of population stratification on genetic analysis are well controlled (10). The Manhattan plots and quantile-quantile plots for male/female and North/South datasets indicate that there is no inflation of the associations (SM Figs 1-4).

We conducted the following two-stage consecutive analyses:

Stage I. Sex-stratified GWAS to identify sex-specific loci and pathways associated with longevity and make international comparisons, including three steps: a) Single SNP analysis using sex-specific independent datasets from North and South regions of China as discovery and evaluation samples; b) Comparisons with the results from the longevity GWAS of the European longevity genetics consortium (IDEAL) and the U.S. New England centenarian study (NECS); c) Sex-specific pathway analysis.

Stage II. Polygenic risk score (PRS) analyses to further assess the joint effects of the groups of sex-specific loci identified in Stage I and explore the sex differences in genetic association with longevity, including three steps: a) PRS analyses for males and females separately (methods section M2.1); b) Integrated PRS analysis, based on estimates of the odds ratios of longevity among those with different combinations of sex and the genotype (methods section M2.2 and SM section S4); c) Estimates of the sex-specific genetic relative benefit/loss ratios (methods section M2.3).

Single SNP analysis for males and females separately

The recent literature on GWAS indicates that researchers defined *a priori*

discovery threshold of $P < 10^{-5}$, $P < 10^{-4}$, or $P < 10^{-3}$, depending on the circumstances of the research topics and the characteristics of the datasets used (14-16). In the present study, we aim to identify SNPs which may individually have very small effects, but may jointly have large effects on the complex trait of longevity in men and women, respectively. Thus, we believe that it is reasonable to choose a modest *a priori* threshold of $P < 10^{-3}$ for the sex-stratified discovery stage.

As shown in Table 1, we identified 11 independent loci ($r^2 < 0.1$, representing 24 SNPs) associated with longevity that replicated in male discovery and evaluation datasets of North and South regions of China and reached $P < 10^{-5}$ in the male North-South combined dataset. However, they were not significant ($P > 0.05$) in the female North-South combined dataset.

Of the 11 loci that replicated in males, rs1950902 in the *MTHFD1* gene is a non-synonymous SNP that causes a C to T transition at nucleotide 401 resulting in an arginine to lysine substitution at amino acid 134 (C401T; R134K). A previous study showed that *MTHFD1* 401 CC was associated with a significantly increased risk of gastric cancer (adjusted OR, 1.43; 95% CI, 1.14-1.80) compared with 401 TT/TC genotypes (17). In the present study, the *MTHFD1* 401 T allele is highly associated with longevity in males ($P = 1.09 \times 10^{-7}$), but was not significant in females ($P = 0.949$) (Table 1). A locus (rs12199884) in *PKHD1* (polycystic kidney and hepatic disease-1) is negatively associated with longevity in males ($P = 4.11 \times 10^{-6}$) but not significantly associated with longevity in females ($P = 0.941$).

--Table 1 here--

Table 2 presents the 12 independent loci (representing 38 SNPs), whose associations with longevity were replicated in the female discovery and evaluation datasets of the North and South regions of China and reached $P < 10^{-5}$ in the female North-South combined dataset. None of these loci is significant ($P > 0.05$) in males. Of these 12 loci, SNP rs71352238 in the promotor region of *TOMM40* is associated with aging in females ($P = 9.99 \times 10^{-6}$) but not in males ($P = 0.79$) in a study involving 5 UK

cohorts of older adults (3511 individuals born between 1921 and 1936) and 3 Swedish replication cohorts (N=1367) (18). Previously, two SNPs (rs17047650, rs10433502) at *FAM19A1* were discovered to have protective effects on aging ($P < 5.0 \times 10^{-5}$) in a GWAS of 1,385 subjects (19). However, in the present study these 2 SNPs showed no associations with longevity and no sex difference; instead, another SNP rs1027238 at *FAM19A1* was associated with longevity in females ($P = 2.79 \times 10^{-6}$) but not in males ($P = 0.374$). SNPs in/near *TBX3* are significantly associated with longevity in females, but not in males in the present study, which is consistent with previous findings that *TBX3* played an important role in mammary gland development and breast cancer with a close relationship to estrogen (20).

In addition to the 11 male- and 12 female-specific replicated top loci ($P < 10^{-5}$ in one sex but $P > 0.05$ in other sex), our sex-stratified GWAS found that 71 male- and 98 female-specific independent loci (representing 486 male and 358 female SNPs, respectively) are associated with longevity in one sex ($10^{-5} \leq P < 10^{-4}$), but are not significant in the other sex ($P > 0.05$). We also found that 607 male- and 783 female-specific independent loci (representing 4,551 male and 5,450 female SNPs, respectively) are associated with longevity in one sex ($10^{-4} \leq P < 10^{-3}$), but not significant in the other sex ($P > 0.05$).

Comparison of Han Chinese sex-specific longevity loci with results of European and U.S. GWAS on longevity

We performed a comparative analysis in which we used the Han Chinese sex-stratified GWAS as discovery and two European/American sex-stratified GWAS on longevity (IDEAL and NECS) as evaluations. The IDEAL had 1,865 males aged 85+ and 5,400 females aged 85+ as cases and 16,121 controls aged less than 65, drawn from 14 studies in the Netherlands, Denmark, Iceland, Germany, Italy, the United Kingdom and Sweden (7). The NECS had 339 male and 864 female centenarians (mean age 103.3) as cases and 3,946 middle-aged controls (mean age 36.4) (2).

Among the 904 sex-specific loci in the Han Chinese GWAS which are associated with longevity at a suggestive significance level ($P < 10^{-4}$) in one sex but not significant in the other sex ($P > 0.05$), 321 SNPs were available in the IDEAL GWAS and 722 SNPs were available in the NECS GWAS; these available SNPs were used for the comparison analysis. The results in SM Table 2 show six independent sex-specific loci associated with longevity ($P < 10^{-4}$) in the Han Chinese GWAS that replicated with at least nominal significance ($P < 0.05$) in one sex but were not significant in the other sex in IDEAL GWAS (2 loci) or in NECS GWAS (4 loci). Among these six loci, the rs60210535 of *LINC00871* replicated well between Chinese females ($P = 4.6 \times 10^{-5}$) and U.S. females ($P = 9.0 \times 10^{-5}$), but is not significant in both Chinese and U.S. males ($P > 0.05$). Another female-specific independent locus *ABCG2* has a $P = 6.8 \times 10^{-5}$ in CLHLS females and a $P = 0.003$ in IDEAL females but is not significant ($P > 0.05$) in both Chinese and European males. *ABCG2* is expressed in multiple tissues, participates in xenobiotic metabolism, and is a well-known breast cancer resistance protein (BCRP) (21). Hence, *ABCG2/BCRP* may contribute to the natural resistance and longevity of normal stem cells. *LINC00871* is a non-coding RNA gene and its function is uncertain.

Sex-specific pathway analysis

The results of our sex-specific pathway analysis (Methods section M1.2) suggest that there are sex-specific differences in the biochemical pathways that influence human longevity. In males, 11 pathways were significantly ($P < 0.005$ and $FDR < 0.05$) enriched and associated with longevity (SM Table 3). These pathways are mainly enriched for immune and inflammatory response pathways including inflammatory cytokines and Toll-like receptor (TLR) signaling pathways.

In females, 34 pathways were significantly enriched ($P < 0.005$ and $FDR < 0.05$) and clustered to metabolic pathways (SM Table 4). The tryptophan metabolic pathway and the PPAR γ coactivator-1 α (PGC-1 α) pathway were among the top pathways in

this set.

The PRS analyses for males and females separately

There are three main reasons to conduct PRS analyses following the sex-stratified GWAS. First, as presented earlier, each of the sex-specific loci associated with longevity identified in our sex-stratified GWAS has a very small effect, none of them reached genome-wide significance, and further assessments of their joint effects are needed. Second, although the identified sex-specific loci individually have no significant effect in the other sex ($P > 0.05$), their joint effects could be significantly associated with longevity in the other sex (see SM section S5 for numerical illustrations). Thus, we must evaluate this possibility through PRS analysis with one sex as discovery and other sex as target to further reconfirm or disapprove the candidate sex-specific loci identified in our sex-stratified GWAS. Third, in general, it is impossible to address the present study's main research question on whether the genetic association with longevity is stronger in males or females to better understand the *male-female health-survival paradox* (1), solely based on the single SNP analysis of the sex-stratified GWAS. We therefore must follow up to conduct PRS analysis to explore the sex differences in the joint effects of the identified loci. This is consistent with the final conclusion of Rietveld et al. (22) "Identifying SNPs and constructing polygenic scores are steps toward usefully incorporating genetic data into social-science research." The methods used for our PRS analyses are described in Method section M2.

We first conducted PRS analysis for males and females separately, using the dataset of one sex as the discovery sample and the independent dataset of other sex as the target sample (Method section M2.1). We used the following criterion to judge if a group of loci identified in our sex-stratified GWAS represent true sex-specific longevity loci: if they have joint effects associated with longevity and reach a significance level of $P < 10^{-8}$ in one sex but are not jointly significant in the other sex

($P > 0.05$), they are reconfirmed as sex-specific longevity loci; otherwise, they are not. Note that the lack of significance in the other sex in our PRS analyses is not due to lack of statistical power, because the power of our PRS analyses is excellent for both sexes (to be discussed later).

Panel (A) of SM Table 5 shows that the 11 male-specific independent loci have joint effects associated with longevity and had $P = 1.6 \times 10^{-14} \sim P = 1.1 \times 10^{-10}$ in the male North-South combined dataset. However, as shown in panel (B) of SM Table 5, the joint effects of the 11 male-specific loci are not associated with longevity in females ($P > 0.05$); thus, they are confirmed to be male-specific longevity loci. Similarly, the results presented in panels (A) and (B) of SM Table 6 indicate that the joint effects of the 12 female-specific loci are associated with longevity and exceed the significance level of $P < 10^{-8}$ in females, and that these 12 female-specific loci's joint effects are not associated with longevity in males ($P > 0.05$).

The results of the foregoing PRS analyses of the 11 male- and 12 female-specific loci reconfirmed that they are truly sex-specific longevity top loci. However, the question remains: Could we identify larger sets of loci that fulfill the criterion of being a group of sex-specific longevity loci, jointly significant with $P < 10^{-8}$ in one sex but not jointly significant in other sex ($P > 0.05$), with relaxed P-value thresholds such as $P < 10^{-4}$ or $P < 10^{-3}$ in selecting the loci to be included for constructing the PRS, that also have been used in the literature (14-16,23-24)? To address this question, as described in SM section S5, we further identified four additional exclusive groups of sex-specific longevity-associated loci that fulfill the criterion, following a trial and error approach for selecting an ideal P_T (P-threshold) to provide the best-fitting PRS using the PRSice method and software (25). The four additional groups of sex-specific longevity loci are: i) 44 male-specific longevity strong loci with a $10^{-5} \leq P < 10^{-4}$ in males but $P > 0.4$ in females; ii) 58 female-specific longevity strong loci with a $10^{-5} \leq P < 10^{-4}$ in females but $P > 0.35$ in males; iii) 191 male-specific longevity moderate loci with a $10^{-4} \leq P < 10^{-3}$ in males but $P > 0.75$ in females; iv) 311 female-specific

longevity moderate loci with a $10^{-4} \leq P < 10^{-3}$ in females but $P > 0.7$ in males.

We conducted additional PRS analyses on these four groups of loci. The results of these PRS analyses presented in panels (A) and (B) of SM Tables 7-10 indicate that the 44 male- and 58 female-specific strong loci ($10^{-5} \leq P < 10^{-4}$ in one sex, $P > 0.4$ or $P > 0.35$ in other sex) and the 191 male- and 311 female-specific moderate loci ($10^{-4} \leq P < 10^{-3}$ in one sex, $P > 0.75$ or $P > 0.7$ in other sex) are jointly associated with longevity with a $P < 10^{-8}$ in one sex but are not jointly significant in the other sex ($P > 0.05$). Note that the extremely low P values of the joint effects in some of the sex-specific PRS analysis (Figs 2-3 and SM Tables 7-10) are not unique, as some other published PRS studies also reported such extremely small p-values of the joint effects (26).

We decided not to use the much larger number of loci with a $P < 0.01$ or $P < 0.05$ in one sex but $P > P_T$ in the other sex in our PRS analyses because it could lead to overfitting -- adding complexity but increasing bias (27). We did not find any loci associated with longevity with a $P < 10^{-3}$ in both sexes but in opposite directions. We found 51 loci with a $10^{-3} \leq P < 0.01$ in both sexes but in opposite directions and 1,171 loci with a $10^{-3} \leq P < 0.05$ in both sexes but in opposite directions. We do not include these loci in our present PRS analyses because they do not meet our *a priori* discovery threshold ($P < 10^{-3}$) and due to space limitation, but they may be analyzed in further investigations.

Integrated PRS analysis on sex-specific longevity loci to better understand sex differences

To quantify and better understand the sex differentials in genetic associations with longevity, we conducted integrated PRS analyses to assess differences in the odds ratio of longevity between those who have different combinations of the statuses of sex (male, female) and the genotypes measured by the trisections of the PRS (high, medium, low) summarizing the identified sex-specific longevity loci (Methods section

M2.2 and SM section E4). As shown in Fig 1a, compared to males with the Middle-PRS as a reference group, the likelihood of longevity (measured by becoming a centenarian) of males with Low-PRS is 69.3% smaller ($P=6.1 \times 10^{-14}$) and of males with High-PRS is 152.1% larger ($P=4.3 \times 10^{-11}$). However, there is little difference in the odds ratios of longevity between females who have Low-PRS, Middle-PRS or High-PRS of the 11 male-specific longevity loci, and the curve is very flat (i.e., no effects in females).

Fig 1b demonstrated that, compared with females with Middle-PRS as a reference group, the likelihood of longevity of women with Low-PRS is 46.6% smaller ($P=3.6 \times 10^{-12}$), and for women with High-PRS is 95.5% larger ($P=1.9 \times 10^{-13}$). But there is little difference in odds ratios of longevity among males who have Low-PRS, Middle-PRS or High-PRS of the 12 female-specific longevity loci and the curve is very flat (i.e., no effects in males).

The results of the integrated sex-specific PRS analyses through cross-sex comparisons within each of the genotypes of Low-PRS and High-PRS presented in Figs 2-3 for the 44 male- and 58 female-specific longevity strong loci ($10^{-5} \leq P < 10^{-4}$ in one sex, $P > 0.4$ or $P > 0.35$ in other sex) and 191 male- and 311 female-specific longevity moderate loci ($10^{-4} \leq P < 10^{-3}$ in one sex, $P > 0.75$ or $P > 0.7$ in other sex) all show the same patterns as depicted in Fig 1.

--Figs 1-3 and Tables 3-5 about here---

Sex-specific genetic relative benefit/loss ratios

Results of the own-sex-specific longevity loci

As indicated in Tables 3-5, we estimated cross-sex relative differences of the odds ratios of longevity of different combinations of sex and the genotypes measured by trisections of sex-specific PRS: Males (M) vs. females (F) from males' perspective, i.e. (M-F) / F; and Females vs. males from females' perspective, i.e. (F-M) / M. The positive or negative cross-sex relative differences represent the sex-specific relative

benefit or relative loss (compared to the other sex) due to carrying the genotypes. More specifically, men who carry the High-PRS (or Low-PRS) of the male-specific (i.e. own-sex) longevity loci would have relative benefit (or relative loss) compared to women, because the High-PRS (or Low-PRS) of male-specific longevity loci have significant positive (or negative) effects in men but no or little effects in women (the male curve is sharp but the female curve is very flat, see Figs 1a, 2a and 3a). Similarly, women who carry the High-PRS (or Low-PRS) of the female-specific (i.e. own-sex) longevity loci would have relative benefit (or relative loss) compared to men, as the effects of the female-specific longevity loci are significant in females but have no or little effects in men (the female curve is sharp but the male curve is very flat, see Figs 1b, 2b and 3b). As described in Methods section M2.3, to quantify the comparisons of the relative benefit and relative loss of the effects of the own-sex-specific longevity loci compared to the other sex, we define the *sex-specific genetic relative benefit/loss ratio for the own-sex-specific longevity loci* as: the ratio of absolute values of the cross-sex relative differences in the odds ratios of longevity of carrying the genotype of High-PRS versus Low-PRS based on the own-sex-specific loci (see SM Table 13 for the formulas).

As shown in part (a) of panel (i) in Table 3, for the 11 male-specific longevity loci, the odds ratio of longevity in males with the High-PRS is 72.6% higher (benefit) than that of females with the same High-PRS; and the odds ratio of longevity in males with the Low-PRS is 76.5% lower (loss) than that of females with the same Low-PRS. Thus, the male genetic relative benefit/loss ratio (compared to females) due to carrying the High-PRS or Low-PRS of the 11 male-specific (own-sex) longevity loci is 0.95 ($= 72.6\% / |-76.5\%|$). Similar comparisons shown in part (b) of panel (i) in Table 3, however, indicate that the female genetic relative benefit/loss ratio (compared to males) due to the 12 female-specific (own-sex) longevity loci is 6.65 ($= 165.6\% / |-24.9\%|$), which is 7.0 times as high as the male genetic relative benefit/loss ratio due to the 11 male-specific longevity loci.

The estimates presented in panel (I) of Tables 4-5 for the sex-specific longevity strong loci or moderate loci reveal the same pattern as shown in panel (I) of Table 3, namely, the female genetic relative benefit/loss ratios due to own-sex-specific longevity loci are substantially higher than that of males (Tables 4-5).

Results of the other-sex-specific longevity loci

As shown in Fig 1a, compared to males, females who carry the High-PRS (or Low-PRS) of the male-specific (other sex) longevity loci would have relative loss (or relative benefit), because females have no or little gains from carrying the High-PRS of male-specific longevity loci and they avoid the loss from carrying the Low-PRS of male-specific longevity loci. Similarly, as shown in Fig 1b, men who carry the High-PRS (or Low-PRS) of the female-specific (other sex) longevity loci would have relative loss (or relative benefit) compared to women. As described in Methods section M2.3, to quantify the comparisons of the relative benefits and relative loss of the effects of the other-sex-specific longevity loci, we define the *sex-specific genetic relative benefit/loss ratio for the other-sex-specific longevity loci* as: the ratio of absolute values of the cross-sex relative differences in odds ratios of longevity of carrying the genotype of Low-PRS versus High-PRS based on the other-sex-specific loci (see SM Table 13 for the formulas).

As shown in part (b) of panel (II) in Table 3, for the 12 female-specific longevity loci, the odds ratio of longevity in males with Low-PRS is 33.1% higher (benefit) than that of females with the same Low-PRS, and the odds ratio of longevity in males with High-PRS is 62.4% lower (loss) than that of females with the same High-PRS. Thus, the male genetic relative benefit/loss ratio due to the 12 female-specific (other-sex) longevity loci is 0.53 ($=33.1\% / |-62.4\%|$). However, similar comparisons shown in part (a) of panel (II) in Table 3 indicate that the female genetic relative benefit/loss ratio (compare to males) due to the 11 male-specific (other-sex) longevity loci is 7.74 ($=325.8\% / |-42.1\%|$), which is 14.6 times as high as the male genetic relative benefit/loss ratio due to the 12 female-specific longevity loci.

The estimates presented in panel (II) of Tables 4-5 for the sex-specific longevity strong loci or moderate loci reveal the same pattern as panel (II) of Table 3, namely, the female genetic relative benefit/loss ratios due to other-sex-specific longevity loci are substantially higher than that of males (Tables 4-5).

In sum, the results presented in Figs 1-3 and Tables 3-5 and summarized above indicate that, for the own-sex-specific longevity loci, the ratios of relative benefits due to carrying the High-PRS genotype (gaining the positive effects) to relative loss due to carrying the Low-PRS genotype (suffering the negative effects) are much higher in females than in males. For the other-sex-specific longevity loci, the ratios of relative benefits due to carrying the Low-PRS genotype (avoiding the negative effects) to relative loss due to carrying the High-PRS genotype (missing the positive effects) are also much higher in females than in males.

Integrated PRS analyses on all of the loci associated with longevity in males or females

As outlined above and presented in Figs 1-3 and Tables 3-5, our integrated PRS analyses based on the sex-specific loci have demonstrated that the genetic association with longevity is substantially stronger in females than in males. However, this is not yet a completed story, because genetic associations with longevity in males and females are affected not only by sex-specific loci but also by the other not-sex-specific loci. To obtain a complete picture of sex differences in genetic association with longevity, we further identified the following three groups of independent loci associated with longevity including all of the sex-specific and not-sex-specific loci (SM Table 11):

- (1) 32 loci associated with longevity with $P < 10^{-5}$ in males or females (including 13 loci in males and 19 loci in females);
- (2) 207 loci associated with longevity with $P < 10^{-4}$ in males or females (including 88 loci in males and 119 loci in females);

(3) 1,665 loci associated with longevity with $P < 10^{-3}$ in males or females (including 733 loci in males and 933 loci in females).

We conducted integrated PRS analysis based on these three groups of loci associated with longevity including all of the sex-specific and not-sex-specific loci, and the results indicate that all of the odds ratios of longevity associated with carrying the High-PRS, Mid-PRS or Low-PRS of the loci with $P < 10^{-5}$, $P < 10^{-4}$ or $P < 10^{-3}$ in females are substantially higher than that in males (SM Figs 5-7), except the estimates for the Low-PRS of the loci with $P < 10^{-3}$ are not significant and show no sex difference (SM Fig 7).

DISCUSSION

Our sex-stratified GWAS identified 11 male-specific and 12 female-specific independent loci associated with longevity ($P < 10^{-5}$), and replicated in independent sex-stratified North and South China datasets. Notably, all of these newly identified and replicated sex-specific loci were not significant ($P > 0.05$) in the other sex. We found that the rs60210535 of *LINC00871* overlapped between Chinese females ($P = 4.6 \times 10^{-5}$) and U.S. females ($P = 9.0 \times 10^{-5}$), but were not significant in both Chinese and U.S. males ($P > 0.05$).

Our PRS analyses showed that each of the six exclusive groups of sex-specific loci of 11/12 male/female top loci ($P < 10^{-5}$ in one sex, $P > 0.05$ in other sex), 44/58 male/female strong loci ($10^{-5} \leq P < 10^{-4}$ in one sex, $P > 0.4$ or $P > 0.35$ in other sex) and 191/311 male/female moderate loci ($10^{-4} \leq P < 10^{-3}$ in one sex, $P > 0.75$ or $P > 0.7$ in other sex) are jointly and significantly associated with longevity with a $P < 10^{-8}$ in one sex, but not jointly associated with longevity in the other sex ($P > 0.05$).

In our Single SNP analysis and PRS analysis, some loci that are highly associated with longevity in one sex are no longer statistically significant ($P > 0.05$) or become only marginally significant with substantially increased P -value when the sexes-combined dataset is used (adjusted for sex as a covariate), as was done in all

previously published GWAS on longevity (2,7,9,10,28). For example, among the 11 male- and 12 female-specific top loci, two male top loci are not significant any more ($P > 0.05$) and P -values of all of the other sex-specific top loci increase substantially in the sexes-combined analysis (panel (C) of Tables 1 and 2). These results imply that, if $P < 10^{-5}$ is used as the threshold for the North-South combined dataset, all of the 11 male- and 12 female-specific top loci would not be significant in the sexes-combined analysis adjusted for sex as a covariate, following the classic approach of GWAS on longevity. Note that male-female combined analyses presented in panel (C) of Tables 1-2 and SM Tables 5-10 have much larger sample sizes but the P -values increased substantially, compared to the sex-specific analyses (panels (A) of Tables 1-2 and SM Tables 5-10). This is because the associations of the sex-specific loci with longevity are substantially offset by non-significance in the other sex, which implies that previously published GWAS on longevity identified sex-independent genetic variants (such as *APOE*, *5q33.3*, *IL6* and *FOXO3A*) but missed sex-specific longevity loci. This is consistent with the conclusion of Ober et al. (29) that “Genetic studies that ignore sex-specific effects in their design and interpretation could fail to identify a significant proportion of the genes that contribute to risk for complex diseases”.

Our pathway analysis revealed that inflammatory cytokines and Toll-like receptor (TLR) signaling pathways are enriched in 11 of identified male-specific pathways. Clinical data demonstrate that males and females differ regarding their innate, humoral and cell-mediated responses to viral challenge (30). For example, males develop lower antibody responses and show significantly lower vaccine efficacy than females. Moreover, it is well known that longevity is associated with sex-specific differences in the immune system, and that there is a progressive decline in immunity and dysregulated inflammatory response in males (31-32). Consistent with these trends, and with previous genetics findings (33-34), we found that the pro-inflammatory cytokine interleukin 6 (IL-6) pathway is significantly associated with

longevity in males. Furthermore, we identified the signaling pathway for the pathogen recognition receptor TLR3 as the most significant pathway associated with male longevity. Others previously reported that the TLR3 signaling pathway is dysregulated in elderly humans (35). TLR3 signaling evokes IL-6 production (36), and initiates innate immunity and facilitates adaptive immunity by promoting maturation of dendritic cells (36,37). It is plausible to hypothesize that altered increasing dysregulation of the IL-6 and TLR3 signaling pathways, that may happen normally, renders males more susceptible to bacterial and viral infections than females; conversely, in long-lived males, altered IL-6 and TLR3 signaling pathways may provide greater protection against these challenges.(38)

Thirty-four female-specific longevity pathways are enriched for the tryptophan metabolic and PGC-1 α pathways. Our findings regarding tryptophan are supported by the recently reported metabolic signature of extreme longevity in North-Italian centenarians comprised mostly of females (113 females out of 143 subjects, mean age 100.9 \pm 2.1) (39). The tryptophan levels in blood serum were significantly lower ($p < 0.001$) in female centenarians (70.3 \pm 11.6) compared with elderly females (79.7 \pm 11.9) but the differences was not so significant ($p > 0.01$) in male centenarians (74.9 \pm 10.2) compared with elderly males (81.7 \pm 13.9). Tryptophan metabolism participates in a number of key processes, ranging from regulating innate and adaptive immunity (40) to supporting intermediary metabolism via the provision of NAD⁺ and NADP, to the biosynthesis of serotonin and related signaling molecules. PGC-1 α is the master regulator of mitochondrial biogenesis and function because it promotes the expression of many of the >1000 nuclear genes that encode mitochondrial proteins, and also participates in the regulation of innate immunity (41). One product of tryptophan metabolism, NAD⁺, is also a cofactor for sirtuins, which have been implicated in inflammation and stress resistance and aging. Coincidentally, Sirtuin 1 (*SIRT1*) deacetylates PGC-1 α and enhances PGC-1 α activity. (42). Aging is associated with progressive mitochondrial dysfunction, and while the ultimate cause

for this dysfunction is unknown, insufficient NAD⁺ availability and SIRT1 enzymatic may be contributing factors (43). In considering the female and male longevity-associated pathways together, the potential involvement of the tryptophan and PGC-1a pathways in the regulation of the immune system, suggest that females and males have optimized different approaches for solving the same biological riddle.

The results of the integrated PRS analyses on the genetic associations with longevity presented in Figs 1-3 and SM Figs 5-7 are all in favor of females. Could these results be driven by the fact that, on average, female cases and controls live longer than their male counterparts? Our answer is no because our sex-stratified GWAS and PRS analyses are not based on sex differences in lifespan among cases of centenarians and among middle-aged controls. Instead, we count all male and female centenarians as “cases” disregarding their lifespan difference and all male and female middle-aged individuals as “controls” disregarding their lifespan difference. As explained in SM section S1, we estimate the sex-specific odds ratios (and *P* values) of longevity associated with carrying the genotypes of High-PRS, Mid-PRS or Low-PRS by using logistic regression to compare the sex-specific frequencies of the genotypes between the centenarians and the middle-aged controls, with no effects by how long each of the male and female centenarians and controls would survive. Clearly, the sex-specific odds ratios of a genotype’s association with longevity in Figs 1-3 and SM Figs 5-7 are determined by the effects of the genotypes, and they are not driven by the fact that on average women live longer than men.

Could the stronger genetic association in females than in males be driven by the larger sample size of female centenarians compared to males, which is common in all studies of centenarians and in the real world? Our answer is no, mainly for two reasons. First, in China in the 1990s there were 2.3 male centenarians per one million males and 7.8 female centenarians per one million females; this implies that male centenarians are much more stringently selected “longevity-stars” due to higher death rates in males than in females at younger ages. Consequently, as shown in SM Tables

5-10, the proportions of variance in the male longevity trait jointly explained by the male-specific loci are substantially larger than the variance in the female longevity trait jointly explained by the female-specific loci, although the male centenarians sample size is substantially smaller than that of females. This may explain why our female-specific GWAS has a relatively good power and the power of our male-specific GWAS is also reasonably acceptable (SM section S6; SM Table 14); and why the power of our PRS analyses estimated using the AVENGEME method and its R program (24) are excellent for both sexes: 0.997~0.999 for males and 1.00 for females (SM section S6; SM Table 15). In other words, the male centenarians' feature of being more stringently selected "longevity-stars" may offset the shortage of power due to their much smaller sample size compared to female centenarians. Second, additional analysis based on a randomly selected female centenarians/controls sample that has the same size as the male sample demonstrated the same pattern of sex differences as analyses using the total female and male samples (data not shown but available upon request).

We also believe that our findings of stronger genetic association in females than in males are not biased by the larger number of identified female-specific longevity loci than that of males, based on following evidence. First, the PRS score of a group of loci at a given significance level for each of the individuals does not depend on the number of loci included, because the score is constructed by the sum of the number of risk allele copies of each of the loci multiplied by the log of the odds ratio of the trait and then divided by the total number of loci (23,24). Second, the numbers of loci included are exactly the same for males and females in the integrated PRS analyses on each of the six groups of the male- and female-specific loci presented in Figs 1-3, and they all indicate that the genetic association in females is much stronger than that in males. The results of panel (II) in Tables 3-5 indicate that the female genetic relative benefit/loss ratio based on the smaller number of male-specific longevity loci is 8.4-14.6 times as high as the male genetic relative benefit/loss ratio

based on the larger female-specific longevity loci. Furthermore, the odds ratios of longevity in females who carry the genotype of Mid-PRS summarizing either the smaller number of male-specific longevity loci (Figs 1a, 2a and 3a) or the larger number of female-specific longevity loci (Figs 1b, 2b and 3b) are all substantially higher than that of males who carry the same Mid-PRS genotype. Third, as shown in SM Figs 8-10, the integrated PRS analyses based on exactly the same numbers of male-specific and female-specific loci randomly selected from the list of all female-specific loci showed the same pattern of much stronger genetic association with longevity in females than in males. Our findings are consistent with a previous study (44) which demonstrated that the variance in individual height explained by the genetic variants does not depend on larger or smaller number of randomly selected SNPs.

The findings presented in Figs 1-3 and SM Figs 5-7 and discussed above lead us to propose a hypothesis that females' genetic constitution may be substantially more in favor of longevity than males', which may be one of the reasons why females live significantly longer than males, while their socioeconomic status and health conditions are not as good as males (1). General evolutionary theory indicates that mutation accumulation affects the Darwinian fitness of populations and it is the primary route by which senescence and longevity could evolve (45). The influence of genetic factors on health and behavior is conditioned by social, cultural, institutional, and physical environments in which individuals live, work, and play (46), and thus the sex differences in genetic influence on longevity are affected by different environmental factors and roles males and females play in their life course. Compared to men, women contribute much more to childbearing and offspring care. Recent studies developed nonlinear models of the age-specific force of natural selection for genetic mutation accumulations, primarily focusing on age-specific manifestation of genetic load (47-49); these studies indicate that fertility is the key factor of Darwinian natural selection for genetic mutation accumulations driving population survival and growth in human evolution. In addition, the grandmother hypothesis proposes that human

postmenopausal longevity evolved from grandmothers' help with grandchildren care that increased their daughters' and daughters-in-laws' fertility, health and survival. The longer grandmothers live, the greater help they provide to their children and grandchildren, leading to increased longevity in subsequent generations (48-51). Perls et al. (52) found that women who lived to at least age 100 were four times more likely to have had children in their forties than women who survived only to age 73. Other studies also revealed that late childbearing after ages 35 or 40 is positively and significantly associated with longevity, with genetic construction as one of the explanatory factors (53,54). Perls and Fretts (55) proposed that women's longevity advantage over men may simply be a by-product of genetic forces that maximized the length of time during which women could bear and raise children and assist with grandchildren as well. In short, the relevant literature supports our hypothesis that females' advantages in genetic constitution may be accumulated over human evolution to meet their broad reproductive and offspring-caring duties as mothers and grandmothers.

Based on reviews of 22 studies examining the fifteen historical famines worldwide, it was found that in the event of a severe famine, male excess mortality rate often exceeds female excess mortality rate, in particular for infants (56). The "female survival advantage" also has been detected in the famines in post-war China (1959–1961), Bangladesh, and the Sudan (57,58). Excess age-specific mortality rates were found among males compared to females during the heat wave of summer (59). Why did females suffer substantially less in survival than males when faced with the same negative shocks in environmental changes during the famines and other non-war circumstances, especially given that females' socioeconomic status was on average poorer than males? Females' genetic constitution that favors longevity more than males' may serve as one of the plausible explanations.

In sum, our sex-stratified GWAS and novel bio-demographic analysis using the polygenic risk scores have discovered that sex differences in genetic associations

with longevity are remarkable and that females' genetic constitution favors longevity more than males'. We believe that these new findings may be partially explained by a hypothesis that females' advantages in genetic constitution were accumulated during human evolution to meet their reproductive and offspring-caring duties as mothers and grandmothers, mainly based on review and analysis of the relevant literature.

The present study has several limitations. First, although our sample size of centenarians is about 2.7 times as large as other largest published single GWAS of centenarians (2), our sex-stratified sample sizes of centenarians, especially for males, may still not be large enough. Most likely, this accounts for the fact that none of our identified sex-specific Single SNPs individually reached genome-wide significance. Second, while our analyses using centenarians as cases and middle-aged adults as controls are scientifically justified and widely practiced in the literature (see SM section S1) to identify associations between genetic variants and longevity, we did not investigate effects of gene-environment interactions on longevity of reaching age 100+ in the present study. This is due to the lack of long-term follow-up GWAS data from the members of the same birth cohort at young ages and ages 100+ for any population and the incompatibility of the environmental factors of the centenarian cases and middle-aged controls who were born about 50 years apart. Third, while the association with longevity of sex-specific genetic variants (individually and combined into polygenic risk scores) and sex-specific biological pathways are described in this study, these associations cannot be assumed to cause our novel finding of stronger genetic association in females than in males. However, the key findings of this study, remarkable sex differences and females' genetic constitution that favors longevity more than males', are worthy of further investigations by interdisciplinary collaborative efforts, such as replications from other populations, international meta-analysis with much larger sample size, lab tests and *in silico* functional validations.

METHODS

M1 Sex-stratified GWAS

M1.1 Sex-specific Single SNP analysis

To minimize false-negative and false-positive rates, we applied a novel bi-directional discovery-evaluation strategy (60) (note: we use “discovery-evaluation” in single SNP analysis and “discovery-target” in PRS analysis, following the practice in the literature). This involved analyzing the sex-specific North region GWAS datasets for discovery and the sex-specific South region GWAS datasets for evaluation, then vice versa. In this way, we captured all sex-specific SNPs associated with longevity with a p-value lower than the threshold in the discovery stage and nominal significance in the evaluation stage, and fully utilized the available sex-specific independent GWAS datasets of North and South regions (SM section S3).

We performed single SNP analysis of sex-stratified GWAS using logistic regression as implemented in PLINK (1.06) (61). To minimize the effects of population stratification, we adjusted for the top two eigenvectors, which corrected nearly all of the stratification that can be corrected (62). In the sex-stratified and North-South combined data analysis, we also adjusted for the geographic stratification of the North and South regions.

M1.2 Sex-specific pathway analysis

Pathway enrichment analysis was performed to identify biological pathways associated with the sex-specific loci. First, we applied PRSice (Polygenic Risk Score software) (25) to calculate BEST-FIT-P value cutoffs of SNPs used for pathway analysis. Second, an improved gene set enrichment analysis (GSEA) for GWAS, i-GSEA4GWAS (63), was implemented to map genes to pathways. Pathway gene sets with $P < 0.005$ and false discovery rate $FDR < 0.05$ were regarded as significantly associated with longevity. The sex-specific significant pathways were analyzed independently.

M2. The PRS analyses

As reported in the literature, one of the main types of PRS analyses uses the results of loci associated with a health outcome identified in another independent study with large sample size (such as published consortium or meta-analysis) to construct a PRS and test the association in the population under investigation, which likely has relatively small sample size (e.g. ref. 31). Another major type of PRS analyses consists of a two-stage analytical design: (I) Conduct GWAS (or candidate genes analysis) to identify the SNPs associated with the health outcome; (II) Conduct PRS analysis to assess the joint effects of the identified SNPs based on a discovery dataset, and then test these PRS scores using an independent dataset as target sample (22,27,64).

Given the fact that there was no sex-stratified GWAS on longevity prior to the present study, we employ an extended two-stage study design: Stage I. Sex-stratified GWAS to identify the sex-specific loci associated with longevity; Stage II. PRS analyses to assess the joint effects of the sex-specific loci and better understand the sex differences in genetic association with longevity. As described and justified in the section M2.1 below, we use dataset of one sex as discovery sample and the independent dataset of other sex as target sample.

Following the standard approach used in the literature (23,24), we constructed PRS scores as the sum of the number of risk allele copies of each of the selected loci multiplied by the log of the corresponding odds ratio of longevity and then divided by the total number of the selected loci for each of the centenarians and middle-aged controls. Note that it would be impossible to use continuous scores for integrated PRS analysis on different combinations of the genotype and sex to assess the sex differences (Methods section M2.2). A bisection ordered variable of PRS is too rough and inaccurate and the quartiles of PRS are somewhat complicated for comparisons. Thus, we use trisections of PRS (High-PRS, Middle-PRS and Low-PRS), with Middle-PRS as the reference group.

M2.1 The PRS analyses for males and females separately

Our PRS analyses for males and females separately include the following two sub-steps: (A) Assess the joint effects of the sex-specific longevity loci identified in our sex-stratified GWAS, using the dataset of one sex as discovery sample. (B) Test whether the identified sex-specific loci are jointly associated with longevity in the independent dataset of the other sex as target sample. We did not re-identify the sex-specific loci based on North (or South) sex-specific sub-samples only to construct the PRS and test it in the South (or North) sex-specific sub-samples, mainly for three reasons. First, using the North (or South) sex-specific sub-samples only to select loci for constructing PRS would miss substantial numbers of sex-specific loci due to not fully utilizing all available GWAS data (see SM section S3 for justification and numerical illustration). Second, It would be very hard to understand and interpret the sex differences based on the PRS results of 4 ($=2 \times 2$) sub-datasets of North-South and male-female cross-divisions. Third, the purpose of our PRS analyses is to assess the joint effects of the sex-specific loci identified in the sex-stratified GWAS (each of them may have a very small effect on longevity) and to evaluate the sex difference in genetic association with longevity. There is no need to re-identify sex-specific loci using the North (or South) dataset only and test them in the South (or North) dataset, as it does not serve the purpose of our PRS analyses. Note that the basic design of our PRS analyses for males and females separately using the loci identified in the GWAS is consistent with other PRS studies focusing on sex differences (3).

Normally, the PRS in the target sample are constructed using the log of odds ratio of longevity from the discovery sample as the weights because the target sample is used to replicate the results of association found in the discovery sample, or to predict the trait in the target sample using the results of the discovery sample (23,24). However, the purpose of the PRS analysis on the sex-specific loci using the dataset of one sex as discovery sample and dataset of the other sex as target sample in present study (see panels (A) and (B) in SM Tables 5-10) is not for replication or prediction of

the results from discovery sample. Instead, we aim to assess whether the sex-specific loci, whose joint effects are significantly associated with longevity in the discovery sample of one sex, are significant or not in the other sex dataset as target. In short, we aim to identify the differences across sexes. Thus, for the PRS of the other sex target sample, we must use the odds ratios as weights estimated by its own dataset, namely, the other sex dataset, rather than using the odds ratios from the discovery sample (3). This is because using the log of odds ratio of longevity from sample of one sex (discovery) as the weights in the sample of other sex (target) would largely over-estimate the sex differences in genetic association with longevity.

We use the PRSice software, which is specially designed for PRS analysis (25) to estimate the PRS of the selected longevity loci for each of the centenarians and controls. We employed the PRSice's "clumping" method with default option to select tag-SNPs or "independent loci" by excluding all SNPs with Linkage disequilibrium (LD; $r^2 > 0.1$); only the "independent loci" are used to calculate the PRS. We used STATA 12.0 software to perform the sex-specific case/control association analysis on longevity using the ordered variables (trisections) of the PRS. We also used the PRSice software to conduct the analysis based on the continuous PRS for males and females, which indicated similar general patterns as those based on trisections of PRS presented in the SM Tables 5-10.

M2.2 Integrated sex-specific PRS analysis

It is hard to quantify sex differences in genetic association with longevity solely based on GWAS or PRS analyses for males and females separately (Tables 1-2 and SM Tables 5-10). This is because such separated analyses use different reference groups for males and females respectively, making the odds ratios and other estimates for males and females not fully compatible. Thus, to quantify and better understand the sex differentials in genetic associations with longevity, we conducted integrated PRS analyses to assess differences in the odds ratio of longevity between those who have

different combinations of the statuses of sex (male, female) and the genotypes measured by the trisections of the PRS (high, medium, low) summarizing the identified sex-specific longevity loci. The strength of the integrated PRS analysis, which uses one reference group, is that all of the sex-specific odds ratios and other estimates are fully compatible and can be intuitively depicted in the graphics (Figs 1-3 and SM Figs 5-7). Note that this statistical analysis procedure was applied in previous publications (65-66) and we apply it here to the integrated sex-specific PRS analyses using the STATA 12.0 software (see SM section S4 for a technical note).

M2.3 A novel bio-demographic index of “sex-specific genetic relative benefit/loss ratio”

As presented earlier and demonstrated numerically in Tables 3-5, we estimate the cross-sex relative differences of the odds ratios of longevity by different combinations of sex and the genotypes measured by trisections of sex-specific PRS: (I) Males vs. females, $(M-F) / F$; and (II) Females vs. males, $(F-M) / M$. The positive or negative cross-sex relative differences represent the sex-specific relative benefit or relative loss, compared to the other sex, due to carrying the genotypes. For example, women who carry the High-PRS (or Low-PRS) of the female-specific (own-sex-specific) longevity loci would have relative benefits (or relative loss) compared to men, because the High-PRS (or Low-PRS) of female-specific longevity loci have significant positive (or negative) effects in women but no or little effects in men. On the other hand, women who carry the High-PRS (or Low-PRS) of the male-specific (other-sex-specific) longevity loci would have relative loss (or relative benefit) compared to men, because women have no or little gains from carrying the High-PRS of male-specific longevity loci and they avoid the loss from carrying the Low-PRS male-specific longevity loci.

To quantify the relative positive and relative negative effects of the sex-specific longevity loci compared to the other sex, we proposed a novel bio-demographic index, the sex-specific genetic relative benefit/loss ratio, defined as: (A) *sex-specific genetic*

relative benefit/loss ratio for the own-sex-specific longevity loci, which is the ratio of absolute values of the cross-sex relative differences in odds ratios of High-PRS versus Low-PRS; (B) *sex-specific genetic relative benefit/loss ratio for the other sex longevity loci*, which is the ratio of absolute values of the cross-sex relative differences in odds ratios of Low-PRS versus High-PRS (see SM Table 13 for the formulas). The word “relative” in the index name means comparing to the other sex.

ACKNOWLEDGMENTS

This study is supported by the National Natural Science Foundation of China (71233001, 71490732, 71110107025; YZ), National Key Basic Research Program of China (2013CB530700;XLT,TN,YZ) and NIA/NIH (R01AG023627,YZ; 2P01AG031719, JWV,KC,YZ; P30AG028716,ERH; U19AG023122,TP,PS), EU 7th Framework Program (259679, ES).

AUTHOR CONTRIBUTIONS

Y.Z., J.W.V., C.N. and J.M jointly supervised this work. Y.Z., J.W.V., J.L. and J.X.L. designed and supervised CLHLS which recruited the participants. C.N., X.M.L., R.Y. contributed to sample management and genotyping. H.C., X.M.L., R.Y., E.X, Z.C. and C.N. performed bioinformatics quality controls and the statistical analysis. Y.Z., J.M., H.C., X.M.L. and C.N. wrote the initial draft, with critical revisions and comments by J.W.V., X.T., H.Y., M.L., E.H., E.S., P.S, J.D., T.P, K.C., Q.T., C.F., W.G., J.M.R., A.M.O., K.L., J.X., A.Y. and Y.T. All authors revised and gave final approval of this manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

References

1. Van Oyen H, et al. (2013) Gender differences in healthy life years within the EU: an exploration of the “health–survival” paradox. *Int J Public Health* 58(1):143–155.
2. Sebastiani P, et al. (2012) Genetic signatures of exceptional longevity in humans. *PLoS One* 7(1):e29848.
3. Hughes T, et al. (2012) Analysis of autosomal genes reveals gene–sex

- interactions and higher total genetic risk in men with systemic lupus erythematosus. *Ann Rheum Dis* 2012;694 71:694–699.
4. Gilks WP, Abbott JK, Morrow EH (2014) Sex differences in disease genetics: evidence, evolution, and detection. *Trends Genet* 30(10):453-463.
 5. Mielke MM, Vemuri P, Rocca WA (2014) Clinical epidemiology of Alzheimer's disease: assessing sex and gender differences. *Clin Epidemiol* 6:37-48.
 6. Altmann A, et al. (2014) Sex modifies the APOE-related risk of developing Alzheimer disease. *Ann Neurol* 75(4):563-73.
 7. Deelen, J. et al. (2014) Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum Mol Genet* 23(16): 4420-4432.
 8. Sebastiani P, et al (2017) Four Genome-Wide Association Studies Identify New Extreme Longevity Variants. *J Gerontol A Biol Sci Med Sci*, forthcoming. doi:10.1093/gerona/glx027.
 9. Newman AB, et al. (2010) A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *J Gerontol A Biol Sci Med Sci* 65(5): 478-487.
 10. Zeng Y, et al. (2016) Novel loci and pathways significantly associated with longevity. *Sci Rep* 6, 21243.
 11. Cohen J et al. (2013) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. (Routledge, Abingdon-on-Thames).
 12. Jeune B (1995) In search of the first centenarians. *Exceptional Longevity: from Prehistory to the Present*, eds Jeune B, Vaupel J (Odense Univ Press, Odense), pp 11-24.
 13. Xu S, et al. (2009) Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am J Hum Genet* 85(6): 762-774.
 14. Chen Y, et al. (2014) Common variants near *ABCA1* and in *PMM2* are associated with primary open-angle glaucoma. *Nat Genet* 46(10):1115-1119.
 15. International Multiple Sclerosis Genetics Consortium (IMSGC), et al. (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 45(11):1353-1360.
 16. Saxena R, et al. (2013) Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes* 62(5):1746-1755.
 17. Wang L, et al. (2007) Polymorphisms of MTHFD, plasma homocysteine levels, and risk of gastric cancer in a high-risk Chinese population. *Clin Cancer Res* 13(8):2526-3213.
 18. Davis G, et al. (2014) A genome-wide association study implicates the APOE locus in nonpathological cognitive aging. *Mol Psychiatry* 19(1): 76-87.
 19. Shi H, et al. (2012) Genetic variants influencing human aging from late-onset Alzheimer's disease (LOAD) genome-wide association studies (GWAS). *Neurobiol Aging* 33(8):1849.e5-18.
 20. Fillmore CM et al. (2010) Estrogen expands breast cancer stem-like cells through paracrine FGF/Tbx3 signaling. *Proc Natl Acad Sci U S A* 107(50): 21737-21742.

21. Natarajan K, et al. (2012) Role of breast cancer resistance protein (BCRP/ABCG2) in cancer drug resistance. *Biochem Pharmacol* 15:1084–1103.
22. Rietveld CA, et al. (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340(6139), 1467-1471.
23. Purcell SM, et al. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460(7256): 748-752.
24. Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 9(3): e1003348.
25. Euesden J, Lewis CM, O'Reilly PF (2015) PRSice: polygenic risk score software. *Bioinformatics* 31(9): 1466-1468.
26. Power RA, et al. (2015) Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci* 18(7): 953-955.
27. Levine ME. et al. (2014) A Polygenic Risk Score Associated with Measures of Depressive Symptoms Among Older Adults. *Biodemography Soc Biol* 60:199–211.
28. Broer L, et al. (2015) GWAS of longevity in CHARGE consortium confirms *APOE* and *FOXO3* candidacy. *J Gerontol A Biol Sci Med Sci* 70(1):110–118.
29. Ober C, Loisel DA, Gilad Y (2008) Sex-specific genetic architecture of human disease. *Nat Rev Genet* 9(12): 911-922.
30. Klein SL, et al. (2010) The Xs and Y of immune responses to viral vaccines. *Lancet Infect Dis* 10(5):338-349.
31. Goetzl EJ, et al., Gender specificity of altered human immune cytokine profiles in aging, *FASEB J.* 24 (2010) 3580–3589.
32. Hewagama, et al. (2009) Stronger inflammatory/cytotoxic T-cell response in women identified by microarray analysis. *Genes Immun* 10: 509–516.
33. Marttila S, et al. (2013) Transcriptional analysis reveals gender-specific changes in the aging of the human immune system. *PLoS One* 8(6): e66229.
34. Bonafè M, et al. (2001) A gender-dependent genetic predisposition to produce high levels of IL-6 is detrimental for longevity. *Eur J Immunol* 31(8): 2357–2361.
35. Agrawal A, et al. (2007) Altered innate immune functioning of dendritic cells in elderly humans: a role of phosphoinositide 3-kinase-signaling pathway. *J Immunol* 178: 6912–6922.
36. Melkamu T, et al. (2013). TLR3 activation evokes IL-6 secretion, autocrine regulation of Stat3 signaling and TLR2 expression in human bronchial epithelial cells. *Journal of Cell Communication and Signaling* 7(2):109–118.
37. Varthaman A, et al (2016) TLR3-Induced Maturation of Murine Dendritic Cells Regulates CTL Responses by Modulating PD-L1 Trafficking. *PLoS one* 11(12): e0167057. doi:10.1371/journal.pone.0167057.
38. Kong KF, et al. (2008) Dysregulation of TLR3 impairs the innate immune response to West Nile virus in the elderly. *J Virol* 82(15): 7613-7623.
39. Collino S, et al. (2013) Metabolic signatures of extreme longevity in northern Italian centenarians reveal a complex remodeling of lipids, amino acids, and gut microbiota metabolism. *PLoS One* 8(3): e56564.
40. McGaha TL, et al (2012) Amino acid catabolism: a pivotal regulator of innate and adaptive immunity. *Immunol Rev* 249(1): 135-157.
41. Colegio OR, et al (2014) Functional polarization of tumour-associated

- macrophages by tumour-derived lactic acid. *Nature* 513(7519): 559-563.
42. Cantó C, et al. (2009) AMPK regulates energy expenditure by modulating NAD⁺ metabolism and SIRT1 activity. *Nature* 458(7241):1056-1060.
 43. Gomes AP, et al. (2013) Declining NAD⁽⁺⁾ induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging. *Cell* 155(7):1624-1638.
 44. Yang J, et al. (2010) Common SNPs explain a large proportion of heritability for human height. *Nat Genet* 2010; 42(7): 565–569.
 45. Rose MR (1994) *Evolutionary Biology of Aging*: (Oxford Univ Press, Oxford).
 46. Harris KM, et al. (2013) Social, Behavioral, and Genetic Linkages from Adolescence into Adulthood. *Am J Public Health* 103 (S1): S25-S32.
 47. Wachter KW, Evans SN, Steinsaltz D (2013) The age-specific force of natural selection and biodemographic walls of death. *Proc Natl Acad Sci U S A* 110(25): 10141-10146.
 48. Wachter KW, Steinsaltz D, Evans SN (2014) Evolutionary shaping of demographic schedules. *Proc Natl Acad Sci U S A* 111(Suppl 3): 10846-10853.
 49. Evans SN, Steinsaltz DR, Wachter KW (2013) A Mutation-Selection Model for General Genotypes with Recombination. *Memoir 222 (American Mathematical Society, Providence, RI)*. <https://arxiv.org/abs/q-bio/0609046v3>.
 50. Kim PS, Coxworth JE, Hawkes K (2012) Increased longevity evolves from grandmothing. *Proc Biol Sci* 279(1749): 4880-4884.
 51. Kim PS et al. (2014) Grandmothing drives the evolution of longevity in a probabilistic model. *J Theor Biol* 353: 84-94.
 52. Perls TT, et al. (1997) Middle-aged mothers live longer. *Nature* 389(6647): 133.
 53. Smith KR, Mineau GP, Bean LL (2002) Fertility and post-reproductive longevity, *Social Biology* 49(3-4): 185-205.
 54. Zeng Y, Vaupel JW (2004) Association of Late Childbearing With Healthy Longevity among The Oldest-Old in China. *Population Studies* 58(1): 37-53.
 55. Perls TT, Fretts R (1998) Why women live longer than men. *Scientific American Presents* 9(2):100 103.
 56. Macintyre K (2002) Famine and the female mortality advantage. *Famine demography: Perspectives from the past and present*, eds Allen T, et al. (Oxford Univ Press, Oxford), pp 240-260.
 57. Watkins SC, Menken J (1985) Famines in historical perspective. *Popul Dev Rev* 11(4):647-675.
 58. Ashton B, et al. (1992) Famine in China, 1958–61. *The Population of Modern China*, eds Poston Jr DL, Yaukey D (Springer Science + Business Media, New York), pp 225-271.
 59. Robine JM, Michel JP, Herrmann FR (2012) Excess male mortality and age-specific mortality trajectories under different mortality conditions: A lesson from the heat wave of summer. *Mech Ageing Dev* 133(96): 378–386.
 60. Jia P, et al. (2012) Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia. *PLoS Comput Biol* 8(7):e1002587.
 61. Purcell S, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3): 559-575.

62. Price AL, et al. (2008) Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* 4(1): e236.
63. Zhang K, et al. (2010) i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res* 38(Web Server issue):W90-95.
64. Lee SH, et al. (2012) Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 44(3): 247-250.
65. Zeng Y, et al. (2015) GxE Interactions between FOXO Genotypes and Tea Drinking Are Significantly Associated with Cognitive Disability at Advanced Ages in China. *J Gerontol A Biol Sci Med Sci* 70(4):426–433.
66. Rudolph A, et al. (2015) Investigation of gene-environment interactions between 47 newly identified breast cancer susceptibility loci and environmental risk factors. *Int J Cancer* 136(6): E685-E696.

Table 1. The 11 independent loci significantly associated with longevity and replicated in *male* North and South regions datasets, reached $P < 10^{-5}$ in North-South combined male dataset, but not significant ($P > 0.05$) in female combined dataset

SNP	Chr	Nearby gene	(A) Males bi-directional discovery-replication analysis						(B) Females (North-South combined)			(C) Two sexes (North-South combined)			
			North n=286 cases and 508 controls		South n=278 cases and 265 controls		North-South combined n=564 cases and 773 controls		n=1,614 cases and 1,526 controls			n=2,178 cases and 2,299 controls			
			P	Odds ratio	P	Odds ratio	MAF(case /control)	P	Odds ratio	MAF(case /control)	P	Odds ratio	MAF(case /control)	P	Odds ratio
rs1950902	14	MTHFD1	5.02E-04	1.515	1.44E-04	1.649	0.372/0.261	1.09E-07	1.595	0.321/0.310	0.949	1.004	0.334/0.293	3.62E-03	1.145
rs1157755	12	KCNA5	3.00E-04	2.565	1.32E-03	2.446	0.078/0.032	1.93E-06	2.468	0.057/0.059	0.380	0.907	0.062/0.050	6.89E-02	1.188
rs11136774	8	CSMD1	1.75E-04	1.6	7.58E-03	1.475	0.297/0.209	2.62E-06	1.56	0.248/0.249	0.845	0.988	0.260/0.236	1.55E-02	1.131
rs6453914	6	IMPG1	9.10E-04	1.581	2.87E-03	1.633	0.228/0.158	4.05E-06	1.624	0.191/0.182	0.401	1.057	0.201/0.174	1.62E-03	1.192
rs6740706	2	LRRFIP1	9.43E-07	0.536	3.90E-02	0.734	0.195/0.290	2.29E-07	0.610	0.234/0.249	0.39	0.95	0.224/0.263	3.55E-04	0.837
rs12199884	6	PKHD1	9.78E-04	0.365	9.51E-04	0.464	0.039/0.080	4.11E-06	0.428	0.057/0.055	0.941	0.992	0.052/0.064	4.18E-03	0.767
rs79072042	5	NUDT12	3.70E-05	0.470	4.88E-02	0.678	0.088/0.148	7.15E-06	0.552	0.128/0.132	0.597	0.961	0.117/0.137	5.78E-03	0.837
rs200536623	1	SYDE2	7.38E-05	5.917	1.45E-02	3.553	0.035/0.008	8.83E-06	4.527	0.017/0.015	0.541	1.136	0.022/0.012	1.83E-03	1.726
<i>rs138863</i>	22	<i>BRD1</i>	<i>6.95E-03</i>	<i>0.266</i>	<i>5.43E-04</i>	<i>0.204</i>	<i>0.010/0.039</i>	<i>9.48E-06</i>	<i>0.224</i>	<i>0.021/0.026</i>	<i>0.169</i>	<i>0.792</i>	<i>0.018/0.030</i>	<i>1.34E-04</i>	<i>0.575</i>
<i>rs9894443</i>	17	<i>SLC39A11</i>	<i>2.56E-03</i>	<i>1.388</i>	<i>6.27E-04</i>	<i>1.561</i>	<i>0.416/0.335</i>	<i>8.18E-06</i>	<i>1.447</i>	<i>0.354/0.369</i>	<i>0.406</i>	<i>0.957</i>	<i>0.370/0.358</i>	<i>8.82E-02</i>	<i>1.078</i>
<i>rs73329622</i>	5	<i>STK10</i>	<i>3.29E-03</i>	<i>0.678</i>	<i>9.47E-04</i>	<i>0.600</i>	<i>0.180/0.259</i>	<i>9.23E-06</i>	<i>0.642</i>	<i>0.215/0.218</i>	<i>0.879</i>	<i>0.991</i>	<i>0.206/0.232</i>	<i>9.07E-03</i>	<i>0.872</i>

Notes: (1) The lines without (or with) *Italic front* contain the results from the analysis using male-specific North (or South) region GWAS dataset as discovery and male-specific South (or North) region GWAS dataset as evaluation (SM section S3). (2) Panel (C) is based on male-female combined dataset adjusted for sex as a covariate. (3) As discussed in the SM section S1, the odds ratios in this and other Tables and Figures cannot be interpreted as the size of pure effects of the genotype on longevity.

Table 2. The **12** independent loci significantly associated with longevity and replicated in *female* North and South regions datasets, reached $P < 10^{-5}$ in North-South combined female dataset, but not significant ($P > 0.05$) in male combined datasets

SNP	Chr.	Nearby Gene	(A) Females bi-directional discovery-evaluation analysis						(B) Males (North-South combined)			(C) Two sexes (North-South combined)			
			North n=829 cases and 904 controls		South n=785 cases and 622 controls		North-South combined n=1,614 cases and 1,526 controls		n=564 cases and 773 controls			n=2,178 cases and 2,299 controls			
			P	Odds ratio	P	Odds ratio	MAF(case /control)	P	Odds ratio	MAF(case /control)	P	Odds ratio	MAF(case /control)	P	Odds ratio
rs12568089	1	ZFP69B	8.05E-04	1.35	1.10E-03	1.353	0.226/0.174	2.69E-06	1.352	0.217/0.204	0.811	1.024	0.224/0.184	6.23E-05	1.237
rs3805586	5	PGGT1B	1.18E-04	1.342	2.36E-02	1.196	0.352/0.295	8.95E-06	1.275	0.314/0.326	0.133	0.878	0.342/0.305	2.70E-03	1.147
rs1027238	3	FAM19A1	4.56E-04	0.636	1.79E-03	0.667	0.073/0.105	2.79E-06	0.652	0.089/0.078	0.374	1.136	0.077/0.096	5.07E-04	0.766
rs12711357	4	FSTL5	1.30E-04	0.736	0.01168	0.786	0.202/0.254	9.09E-06	0.763	0.230/0.241	0.750	0.970	0.209/0.249	7.98E-05	0.818
rs416352	6	NOTCH4	2.45E-03	0.812	5.12E-04	1.323	0.509/0.466	7.81E-06	1.261	0.505/0.484	0.932	1.007	0.505/0.482	2.51E-04	1.173
rs73070152	19	KIR3DX1	1.33E-03	1.439	7.14E-04	1.604	0.115/0.082	8.02E-06	1.477	0.104/0.102	0.975	1.004	0.112/0.089	1.86E-04	1.307
rs13406646	2	CYP1B1-AS1	6.60E-03	1.278	5.07E-04	1.43	0.205/0.163	9.80E-06	1.348	0.185/0.194	0.777	0.972	0.200/0.173	6.12E-04	1.210
rs2161877	12	TBX3	2.92E-03	0.809	3.47E-04	0.747	0.392/0.455	2.71E-06	0.778	0.410/0.417	0.720	0.971	0.397/0.442	5.17E-05	0.834
rs4972778	2	KIAA1715	1.50E-03	0.782	8.44E-04	0.720	0.201/0.257	5.40E-06	0.759	0.241/0.245	0.697	1.036	0.211/0.253	3.04E-04	0.834
rs118113034	6	FRK	2.76E-03	0.408	6.93E-04	0.183	0.007/0.020	8.45E-06	0.320	0.017/0.017	0.896	1.041	0.009/0.019	2.01E-04	0.487
rs12472681	2	LOC1720	1.33E-03	1.908	5.60E-04	2.279	0.046/0.023	5.45E-06	2.004	0.030/0.035	0.544	0.868	0.042/0.027	3.29E-04	1.557
rs71352238	19	TOMM40	4.41E-03	0.692	4.12E-04	0.602	0.066/0.097	9.34E-06	0.654	0.076/0.088	0.296	0.856	0.069/0.094	1.66E-05	0.707

Notes: (1) The lines without (or with) italic front contain the results from the analysis using female-specific North (or South) region GWAS dataset as discovery and female-specific South (or North) region GWAS dataset as evaluation (SM section S3). (2)-(3).The same as in Table 1.

(A note for copy-setting editors: Please locate Tables 3, 4, and 5 immediately underneath Figures 1, 2 and 3, respectively, for readers to easily and intuitively understand the Tables 3, 4 and 5)

Figure 1. Odds ratios of longevity by the combinations of sex and the PRS summarizing the 11 male- and 12 female-specific longevity top loci ($P < 10^{-5}$ in one sex, $P > 0.05$ in other sex)

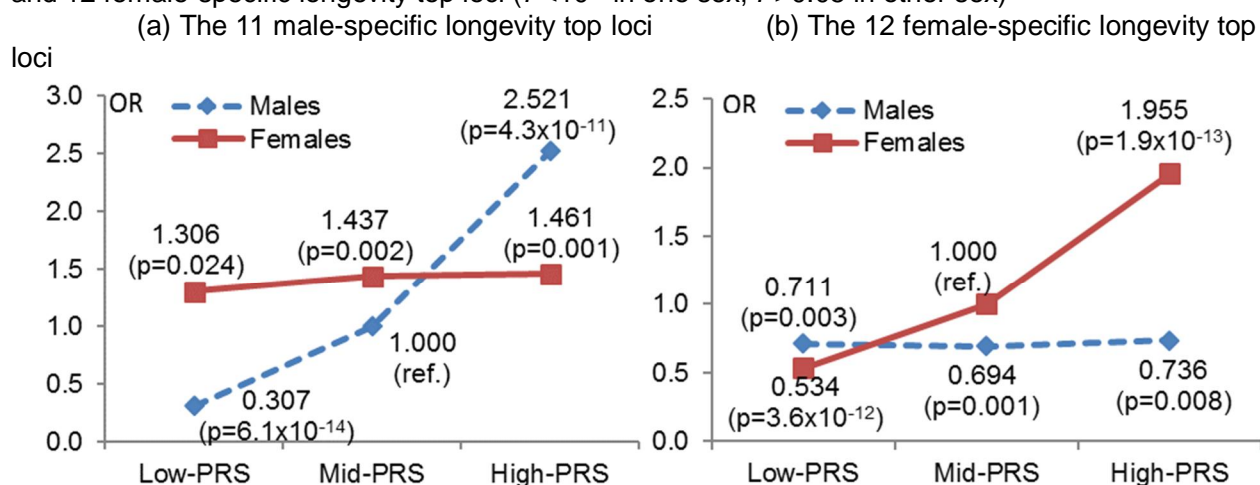


Table 3. Sex-specific genetic relative benefit/loss ratios based on estimates of Figure 1

	a) 11 male-specific longevity top loci			b) 12 female-specific longevity top loci			Female vs Male	
	Low-PRS	Mid-PRS	High-PRS	Low-PRS	Mid-PRS	High-PRS		
(I) Sex diff. Own-Sex longevity loci	(M-F)/F	-76.5%	-30.4%	72.6%	(F-M)/M	-24.9%	44.1%	165.6%
	Male genetic relative benefit/loss ratio: 0.95 (=72.6%/ -76.5%)			Female genetic relative benefit/loss ratio: 6.65 (=165.6%/ -24.9%)			F/M: 7.0 (=6.65/0.95)	
(II) Sex diff. Other-sex longevity loci	Low-PRS	Mid-PRS	High-PRS	Low-PRS	Mid-PRS	High-PRS	Female vs Male	
	(F-M)/M	325.8%	43.7%	-42.1%	(M-F)/F	33.1%		-30.6%
Female genetic relative benefit/loss ratio: 7.74 (=325.8%/ -42.1%)			Male genetic relative benefit/loss ratio: 0.53 (=33.1%/ -62.4%)			F/M: 14.6 (=7.74/0.53)		

Notes: (1) M: Males; F: Females;

Figure 2. Odds Ratios of longevity by the combinations of the sex and the PRS summarizing the 44 male- and 58 female-specific longevity strong loci ($10^{-5} \leq P < 10^{-4}$ in one sex, $P > 0.4$ or $P > 0.35$ in other sex)

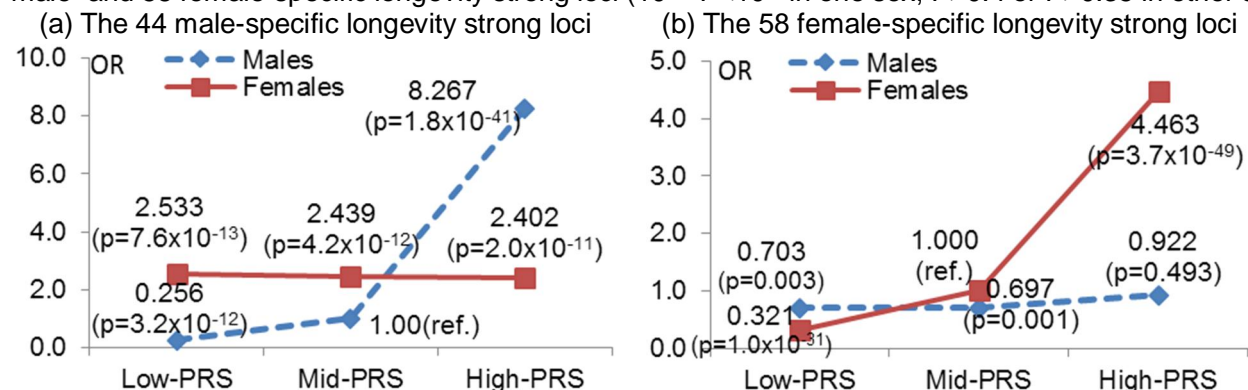


Table 4. Sex-specific genetic relative benefit/loss ratios based on estimates of Figure 2

	a) 44 male-specific longevity strong loci			b) 58 female-specific longevity strong loci			Female vs Male		
		Low-PRS	Mid-PRS	High-PRS		Low-PRS		Mid-PRS	High-PRS
(I) Sex diff. Own-sex longevity loci	(M-F)/F	-89.9%	-59.0%	244.1%	(F-M)/M	-54.4%	43.5%	383.9%	F/M: 2.61 (=7.06/2.71)
	Male genetic relative benefit/loss ratio: 2.71 (=244.1%/ -89.9%)			Female genetic relative benefit/loss ratio: 7.06 (=383.9%/ -54.4%)					
(II) Sex diff. Other-sex longevity loci	(F-M)/M	890.6%	143.9%	-70.9%	(M-F)/F	119.1%	-30.3%	-79.3%	F/M: 8.37 (=12.55/1.50)
	Female genetic relative benefit/loss ratio: 12.55 (=890.6%/ -70.9%)			Male genetic relative benefit/loss ratio: 1.50 (=119.1%/ -79.3%)					

Figure 3. Odds Ratios of longevity by the combinations of the sex and the PRS summarizing the 191 male- and 311 female-specific longevity moderate loci ($10^{-4} \leq P < 10^{-3}$ in one sex, $P > 0.75$ or $P > 0.7$ in other sex)

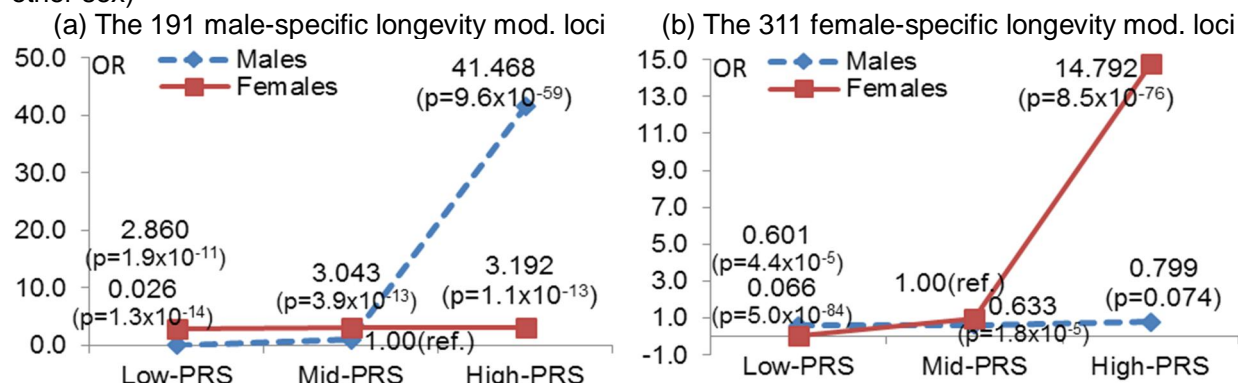


Table 5. Sex-specific genetic relative benefit/loss ratios, based on estimates of Figure 3

	a) 191 male-specific longevity mod. loci			b) 311 female-specific longevity mod. loci			Female vs Male		
		Low-PRS	Mid-PRS	High-PRS		Low-PRS		Mid-PRS	High-PRS
(I) Sex diff. Own-sex longevity loci	(M-F)/F	-99.1%	-67.1%	1199.2%	(F-M)/M	-89.0%	58.0%	1752.4%	F/M: 1.63 (=19.68/12.1)
	Male genetic relative benefit/loss ratio: 12.10 (=1199.2%/ -99.1%)			Female genetic relative benefit/loss ratio: 19.68 (=1752.4%/ -89.0%)					
(II) Sex diff. Other-sex longevity loci	(F-M)/M	10817%	204.3%	-92.3%	(M-F)/F	812.9%	-36.7%	-94.6%	F/M: 13.64 (=117.2/8.59)
	Female genetic relative benefit/loss ratio: 117.2 (=10817%/ -92.3%)			Male genetic relative benefit/loss ratio: 8.59 (=812.9%/ -94.6%)					