SUPPLEMENTAL MATERIALS

Sex differences in genetic associations with longevity in Han Chinese -- Sex-stratified GWAS and polygenic risk score analyses

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 Yashin[j], Angela M O'Rand[j], Liang Sun[k], Ze Yang[k], Wei Tao[l], Anastasia Gurinovich[m], Claudio Franceschi[n], 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, 411005. [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

Table of Contents

Acknowledgements

We are grateful to Xiaomin Shi, Zhaoxue Yin, Jieshi Luo and Yuebin Lv from the China Center for Diseases Control and Prevention; Yuzhi Liu, Chunyuan Zhang and Yun Zhou from Peking University; Zhenyu Xiao, Liqun Tao, Qin Xu and Ye Yuan from the Chinese Center for Aging Science Research; and Zhenzhen Zheng and Jie Zhan from the China Social Science Academy for their contributions to the Chinese Longitudinal Healthy Longevity Survey (CLHLS) and the DNA sample collections; we thank all interviewees and their families for their voluntary participation in the CLHLS study. We thank Xiaopan Liu, Qiuyan Liao, Jingzhi Chen, Hongcheng Zhou and Mingrui Li from the China National GeneBank-Shenzhen for their contributions of sample storage, DNA extractions and genotyping. We are grateful for the thoughtful comments provided by Jessica Sautter and Lene Christiansen.

We are grateful for the following scholars from the EU Longevity Genetics Consortium (including GEHA project consortium) for their contributions to the EU longevity genome-wide association study (GWAS) database which was used for the comparative analysis across Han Chinese, European and U.S. populations in this article: Erik B. van den Akker, Maris Alver, Karen Andersen-Ranberg, Kristin L. Ayers, Marian Beekman, Anna M. Bennet, Vladyslav Bezrukov, Hélené Blanché, Stefan Böhringer, Dorret I. Boomsma, Marie E. Breen, Ruud van der Breggen, Linda Broer, Amke Caliebe, Miriam Capri, Elisa Cevenini, Joanna C. Collerton, Heather J. Cordell, Anton J.M. de Craen, Paolina Crocco, Serena Dato, Karen Davies, Luca Deiana, Cornelia M. van Duijn, Krista Fischer. Friederike Flachsbart, Ian Ford, Pilar Galan, Jutta Gampe, Paolo Garagnani, Eco J.C. de Geus, Efsthatios Gonos, Daníel F. Guðbjartsson, Jennifer Harrow, Diana van Heemst, Bastiaan T. Heijmans, Femke-Anouska Heinsen, Quinta Helmer, Antti Hervonen, Albert Hofman, Jouke-Jan Hottenga, Jeanine J. Houwing-Duistermaat, Bernard Jeune, Palmi V. Jonsson, J.Wouter Jukema, Yoichiro Kamatani, Thomas B.L. Kirkwood, Peter Kristensen, Mark Lathrop, Doris Lechner, Alberta Leon, Patrik Magnusson, Carmen Martin-Ruiz, Susan E. Mcnerlan, Andres Metspalu, Evelin Mihailov, Alberto Montesanto, Simon P. Mooijaart, Anne Murphy, Almut Nebel, Christopher P. Nelson, Ellen A. Nohr, Giuseppe Passarino, Lavinia Paternoster, Pier Giuseppe Pelicci, Markus Perola, Iris Postmus, Michel Poulain, Irene Maeve Rea, Josè Remacle, Fernando Rivadeneira, Jean Marie Robine, Giuseppina Rose, Owen A. Ross, Stefano Salvioli, Nilesh J. Samani, Naveed Sattar, Gary I. Saunders, Stefan Schreiber, Ewa Sikora, Thorkild I.A. Sørensen, Liana Spazzafumo, Hreinn Stefánsson, Kári Stefánsson, David J. Stott, Riin Tamm, Henning Tiemeier, Olivier Toussaint, Stella Trompet, Hae-Won Uh, André G. Uitterlinden, Alexander Viktorin, Rudi G.J. Westendorp, Gonneke Willemsen.

S1. The case-control association analysis

Case-control association analysis of longevity based on logistic regression is often used to estimate associations between genetic variants and longevity by comparing long-lived individuals and middle-aged controls (1-7). This method, which uses long-lived individuals as cases and ethnically/geographically matched middle-aged persons as controls, is based on the fundamental demographic insight that the prevalence of a genetic variant in a population can change with age even though no individuals can change their fixed attribute of the genetic variant; therefore, much can be learned about the impact of the genetic variant on longevity. Ideally, we would use complete cohort data to compare the distributions of genetic variants of members of a cohort at two points of time in

their life span: when they were young versus when some of them reach age 100+, while all of the other cohort members died before age 100. However, complete long-term follow-up GWAS or candidate gene genotype data are not currently available for members of the same birth cohort at young ages and ages 100+ for any population. Thus, previously published association studies of longevity used cross-sectional datasets to compare long-lived individuals (centenarians and/or nonagenarians) as cases and middle-aged individuals as controls, both observed in the same time period. Such cross-sectional case/control analyses are based on two assumptions: (a) The initial distribution of the genetic variants does not differ substantially between the long-lived and middle-aged cohorts, which is reasonable because the basic genetic structure would not change substantially in 40-50 years within the same ethnic population. (b) The basic genetic profiles of the migrants does not differ substantially from non-migrants of the same ethnicity. This assumption is also reasonable, especially for studies of Han Chinese because, as noted in the introduction of the text, China has received remarkably few immigrants in the past. With these two assumptions, one may intuitively understand that the proportion of genetic variants which are positively (or negatively) associated with longevity are significantly higher (or lower) among the centenarians compared to the middle-aged controls, because those who carry the longevity-favoring genetic variants have a better chance of surviving to age 100+, while those with less favorable genetic variants could not reach age 100. The intuitive mechanism based on the two assumptions outlined above has been proven mathematically (8).

Note that the odds ratios of a genotype's association with longevity are usually estimated with PLINK or other standard software using logistic regression based on the differences in proportions of individuals carrying the genotype among cases (centenarians/nonagenarians) and controls (middle-aged) which are determined by the effects of the genotype, gene-environment (GxE) interactions, and other factors related to the genotype. Thus the odds ratios cannot be interpreted as the size of pure effects of the genotype on longevity because the differences in proportions of carrying the genotype in the cases and controls also depend on other factors such as GxE effects.

S2. Samples and data source: the Chinese Longitudinal Healthy Longevity Surveys (CLHLS)

DNA samples and data for the present study are from 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 CLHLS covers approximately 85% of the total population of China. We tried to interview all consented centenarians in the sampled counties and cities. For each centenarian interviewee, we recruited one nearby un-related middle-aged control participant aged 40-59. “Nearby” is loosely defined – it could mean the same village or the same street if available, or in the same town or in the same sampled county or city (9). In the present study, all of the DNA samples from middle-age controls were collected in the same county/city or same province as the nearby centenarians.

Phenotype data were collected in the CLHLS using internationally standardized questionnaires adapted to the Chinese cultural and social context (9). Extensive evaluations of the data quality of the CLHLS, including assessments of mortality rate, proxy use, non-response rate, sample attrition, reliability and validity of major health measures, and the rates of logically inconsistent answers, have shown that the data from the CLHLS surveys are of good quality (10). The genetic samples and data from CLHLS were successfully used in prior published studies on candidate genes and gene-environment interactions relevant to longevity (1, 4, 11-12).

A wide variety of international and Chinese studies (13, 14) have confirmed that age reporting of the Han Chinese oldest-old aged 80+, including centenarians, is reasonably accurate; this is due to the Han Chinese cultural tradition of memorizing one’s date of birth to determine dates of important life events such as engagement, marriage, starting to build a residential house, etc. The accuracy of age reporting in the CLHLS data was reconfirmed by an investigation that compared standard demographic indices of age reporting, such as the single-age distribution of the centenarians, the age progressive ratio among very old adults (e.g., aged 90 years or older), and centenarian density among the oldest-old, between the CLHLS and comparable data from Sweden, Japan, England and Wales,

Australia, Canada, the U.S., and Chile; this study concluded that, although the quality of age reporting of Han Chinese centenarians was not as good as in Sweden, Japan, England and Wales, it was almost as good as in Australia and Canada, slightly better than the average of Whites and Black in the US, and much better than in Chile (15).

Note that the Han Chinese comprise about 93% of the total population in China, with 53 Chinese minority groups comprising 7% of the total population. The sample sizes of any minority group in the CLHLS data are too small for meaningful analysis, so we include Han Chinese samples only in the present study.

S3. A bi-directional discovery-evaluation approach

In the traditional uni-directional discovery-evaluation approach, the entire sample is divided into two datasets, with one dataset used for discovery and the other dataset for replication or evaluation. The top Single nucleotide polymorphisms (SNPs) found in the discovery stage using a pre-determined *P* value threshold are analyzed in the evaluation dataset. The SNPs with a *P* value lower than the threshold (e.g., $P < 1.0 \times 10^{-4}$) in the discovery stage and nominal significance (e.g., $P < 0.05$) in the evaluation stage are identified as significant/replicated SNPs; the significant/replicated SNPs are then examined through association analysis using the discovery-evaluation combined dataset. This traditional uni-directional discovery-evaluation approach is especially useful when the expensive GWAS serves as the first stage of discovery and the much less expensive second stage of evaluation genotypes only the top SNPs with a *P* value lower than the threshold found in the first stage. However, when analyzing two available independent GWAS datasets, such as our present study, the uni-directional discovery-evaluation approach of assigning one GWAS dataset as discovery and another GWAS dataset as evaluation would have a higher false-negative rate, missing a substantial number of significant/replicated SNPs which have a p-value higher than the threshold and lower than the nominal significance level in the discovery GWAS dataset but reach the threshold significance level in the evaluation GWAS dataset.

To avoid the high false-negative rate and to fully utilize all information in the sex-specific independent North and South GWAS datasets available to us, we applied a novel bi-directional discovery-evaluation approach (16) in our single SNPs analysis of sex-stratified GWAS to identify the sex-specific loci significantly associated with longevity. Following this approach, we first analyzed the sex-specific North region GWAS dataset as discovery and the sex-specific South region GWAS dataset as evaluation; we then reversed the process and analyzed the sex-specific South region GWAS dataset as discovery and the sex-specific North region GWAS dataset as evaluation. The results presented in Tables 1 and 2 indicate that our bi-directional approach identified 11 male-specific loci and 12 female-specific loci (23 total sex-specific top loci) significantly associated with longevity and replicated in North and South regions in one sex but not significant in the other sex. The results also show that if we adopted the classic uni-directional approach which fixes North as discovery and South as evaluation without reversing the process, 11 (47.8%) of the 23 sex-specific top loci would be missed because they have a $10^{-3} \leq P < 0.05$ in the North GWAS dataset and $P < 10^{-3}$ in the South GWAS dataset (the 11 loci which would be missed by the classic uni-directional approach are marked with Italic font in Tables 1 and 2). Clearly, our bi-directional approach substantially reduces the false-negative rate.

S4. A technical note on the integrated polygenic risk score (PRS) analysis

If there are significant differences between individuals with different combinations of a genotype and environmental factor (or sex), an interaction between an environmental factor (or sex) and a genotype is present (17). To understand GxE (or G x sex) interaction effects, regressions may be estimated separately for those with different combinations of the genotypes and different categories of the environmental factor (or sex). Sufficiently large sub-sample size is required for each combination if separate regressions are estimated, but this may not be the case in most circumstances, including our present study. Thus, we apply a simple procedure to assess the differences in effects of sex on longevity among those who have different genotypes, without further dividing the samples. Note that

this procedure was applied in previous publications in which the genotype was defined by following the dominant or recessive model and the environmental factors were binary or ordered variables in logistic regression models or Cox proportional hazard models (11,12,18-21). In the present study we extend the procedure to the cases in which genotypes are defined by the ordered variable of trisection categories of the PRS and the binary variable of sex, employing the logistic regression model.

In general, the logistic model may be expressed as:

$$\text{logit}(Y_i) = b_1 G_i + b_2 E_i + b_3 G_i * E_i + \sum_j a_j X_{j_i} \quad (1)$$

where Y_i is the health outcome of the i^{th} individual; G_i represents the genotype (e.g., trisection of PRS) and E_i represents the environmental factor or sex (such as in present study) of the i^{th} individual; X_{ji} is a vector of covariate values corresponding to the i^{th} individual. Coefficients β_1, β_2, β_3 and α_j measure the risk of the health outcome for the corresponding variables.

Let OR_{GE} represent the odds ratio of the health outcome of those with a combination of the genotype status of PRS (G_i) and sex (E_i). Note that there are two major limitations to using the general logistic regression model expressed in Equation (1) and the coefficients β_1, β_2 and β_3 to estimate the odds ratios of health outcome (OR_{GE}) for those individuals with different combinations of the genotype and sex statuses. First, it involves an unrealistic linearity assumption that a one-category change in the ordered independent variable would result in the same effect on the probability of the occurrence of the dependent variable event. For example, using Equation (1) in the present study to estimate OR_{GE} would implicitly assume that the likelihood of longevity for those who carry the High-PRS genotype (G_i=3) is twice as high as that for those who carry the Middle-PRS genotype (G_i=2), and three times as high as that for those who carry the Low-PRS genotype (G_i=1). These assumptions may not be correct in the real world. Second, we have no way to estimate the *P* values of OR_{GE} based on the coefficients β_1, β_2, β_3 and the other outcomes produced by the standard software. To avoid these two major limitations, we estimate OR_{GE} and their *P* values by setting up several exclusive

dummy variables Z_{GE} in the regression equations (22). Z_{GE} represents the combinations of genotype status (which may be binary or ordered variable) and sex (or binary or ordered environmental factor).

In the present study, we estimated the odds ratios of longevity (OR_{GE}) of the combinations of PRS genotypes (the subscript G_i = 1, 2 and 3 for trisections of the Low-PRS, Middle-PRS and High-PRS, respectively) and sex (E_i = 1 for male and 2 for female), using the Equations below:

$$\text{logit}(Y_i) = \sum_{G_i=1 or 3} g_{E_i=1,G_i} Z_{E_i=1,G_i} + \sum_{G_i=1}^{3} g_{E_i=2,G_i} Z_{E_i=2,G_i} + \sum_j a_j X_{ji} \qquad (2)$$

Note that equation (2) modified the conventional linear regression model (1) to an ANOVA (Analysis of Variance)-type model in which the possible interactions among the various combinations of the categorical G and E covariates can be estimated without making linearity assumptions. As shown by the results presented in Figs 1-3 and SM Figs 5-7 based on the logistic regression models expressed in Equations (2) with dummy variables Z_{GE} which represent the combinations of the genotype and sex statuses, the estimates of OR_{GE} adequately and intuitively reveal the sex differences and (G x sex) interaction effects, although there are no explicit interaction terms in the regression equations.

S5. Identification of the sex-specific loci associated with longevity with a more relaxed P_T threshold

Besides the 11 male and 12 female specific longevity top loci ($P<10^{-5}$ in one sex and $P>0.05$ in other sex), our sex-stratified GWAS also identified four additional groups of sex-specific loci associated with longevity: 71/98 male/female loci ($10^{-5} \leq P < 10^{-4}$ in one sex; $P>0.05$ in the other sex) and 607/783 male/female loci ($10^{-4} \leq P < 10^{-3}$ in one sex but $P>0.05$ in the other sex). The results of PRS analysis indicated that these 71/98 male/female loci and 607/783 male/female loci are jointly associated with longevity and jointly reached a significance level of $P<10^{-8}$ in one sex, but they are also jointly and significantly associated with longevity in the other sex ($P=1.1\times10^{-6}$ ~ 2.8×10^{-3} for the 71/98 male/female loci and $P=7.74\times10^{-34}$ ~ 1.88×10^{-17} for the 607/783 male/female loci; data are not

shown but available upon request). Thus these four groups of loci cannot be claimed as sex-specific longevity loci because they are also jointly and significantly associated with longevity in the other sex.

The joint associations with longevity of these four groups of loci in the other sex are because they also include some “not-real-sex-specific loci”, each of which is individually and slightly associated with longevity (P>0.05 and less than a P-threshold) but their combined effects are significant in the other sex. Thus, we need to filter and exclude those “not-real-sex-specific loci” from the 71/98 male/female loci and 607/783 male/female loci in order to identify the groups of true sex-specific longevity loci which fulfill the criterion. Following a trial and error approach for selecting an ideal P_T (P-threshold) to provide the best-fit PRS using the *PRSice* method and software (23), we selected the P_T (P-threshold) to exclude the “not-real-sex-specific loci”. The criteria for selecting the P_T to identify the sex-specific longevity loci is: if a group of loci is jointly associated with longevity and reaches a significance level of $P<10^{-8}$ in one sex but not jointly significant in the other sex (P>0.05), they are sex-specific longevity loci; otherwise, they are not.

After performing the filtering procedure with the criteria described above, we identified the following four additional exclusive groups of sex-specific longevity-associated loci:

(i) 44 male-specific longevity strong loci, with $10^{-5}\leq P<10^{-4}$ in males but P>0.4 in females;

(ii) 58 female-specific longevity strong loci, with $10^{-5}\leq P<10^{-4}$ in females but P>0.35 in males;

(iii) 191 male-specific longevity moderate loci, with $10^{-4}\leq P<10^{-3}$ in males but P>0.75 in females;

(iv) 290 female-specific longevity moderate loci, with $10^{-4}\leq P<10^{-3}$ in females but P>0.7 in males.

We also tried but could not identify much larger sex-specific groups of loci with P<0.05, which are jointly 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). This is likely because inclusion of a very large number of loci with P<0.05 could lead to overfitting -- adding complexity but increasing bias (24).

S6. Power estimates

Using the QUANTO (1.1) software, we evaluated the power of the sex-stratified GWAS to detect genetic effects on longevity for samples of 564 male and 1614 female centenarians and 773 male and 1526 female middle-aged controls to achieve genome-wide significance ($p=5\times10^{-8}$) and suggestive significance ($p=5\times10^{-7}$, $p=5\times10^{-6}$, $p=5\times10^{-5}$, $p=5\times10^{-4}$, $p=5\times10^{-3}$) levels for different minor allele frequency (MAF). We used an additive genetic model and assumed that the prevalence of living to 100 years of age among male and female Han Chinese born in 1898-1908 is $2.33/10^6$ and $7.83/10^6$, respectively, based on estimates from previous study (25) and the ratio of male centenarians to female centenarians from the Chinese census (26). The detailed power estimates results and parameters used are presented in SM Tables 14a-14b and indicate that our female-specific GWAS has a reasonably good power. By comparison, the male-specific GWAS power is not as good as that for the females due to the much smaller sample size of male centenarians, but it is modestly acceptable.

We conducted the power estimates for the sex-specific PRS analyses on longevity using the AVENGEME R program (27, 28). The results and parameters used in the power estimates are presented in SM Table 15 and indicate that the power for both our male-specific and female-specific PRS analyses are excellent: 0.997~0.999 for males and 1.00 for females. Note that there were 2.3 male centenarians per one million males and 7.8 female centenarians per one million females in China in the 1990s (25), which implies that male centenarians may be more stringently selected “longevity-stars” due to higher death rates in males than in females at younger ages. Consequently, the slopes of the regressions for the 11, 44, and 191 male-specific loci among males in Figs 1a, 2a, and 3a are all substantially steeper than the slopes of the regressions for the 12, 58, and 311 female-specific loci among females in Figs 1b, 2b, and 3b. This is a function of the greater selectivity of survival to ages 100+ for the male centenarians than the female centenarians, which implies that there is less random variation around the expected values of the regression equations for the male-specific longevity loci among male centenarians than that for the female-specific longevity loci among female centenarians.

Furthermore, the proportions of variance in the longevity trait jointly explained by the male-specific 11 top loci, 44 strong loci, and 191 moderate loci are all substantially larger than the variance jointly explained by the female-specific 12 top loci, 58 strong loci and 311 moderate loci, respectively (see "Pseudo R^2" in SM Tables 5-10). This may explain why the power of both our male and female PRS analyses estimated using the AVENGEME method and R program (27) are excellent (SM Table 15), because of the larger variance jointly explained by the genetic loci and the higher power of PRS analysis (27). 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.

S7. SM Tables

SM Table 1. Basic characteristics of age structures of the sex-specific CLHLS GWAS datasets

	Males						Females					
	North region		South region		Total		North region		South region		Total	
	Cases: Centen-arians	Mid-age Contr-ols	Cases: Centen-arians	Mid-age Contr-ols	Cases: Centen-arians	Mid-age Contr-ols	Cases: Centen-arians	Mid-age Contr-ols	Cases: Centen-arians	Mid-age Contr-ols	Cases: Centen-arians	Mid-age Contr-ols
n	286	508	278	265	564	773	829	904	785	622	1614	1526
Mean age	101.8	47.9	101.2	50.4	101.5	48.7	103.2	47.1	102.9	50	103.1	48.2
SD	3.24	6.54	3.70	6.64	3.48	6.68	3.44	7.38	3.37	8.05	3.41	7.79

Notes: (1) n--Sub-sample size; SD--Standard deviation; (2) The total number of male and female cases is 2,178 (including 564 male cases and 1,614 female cases); among them, 1,714 cases are centenarians aged 100+ (mean age 103.6) and 464 cases are near-centenarians aged 95-99 (mean age 97.7). To simplify the presentation, we abbreviate them as "centenarians", similar to the abbreviation in the other published centenarians studies (3).

SM Table 2. Sex-specific loci associated with longevity overlapped across the sex-stratified GWAS datasets of Han Chinese CLHLS, European Longevity Consortium (IDEAL) and/or U.S. New England Centenarians Study (NECS)

SNP	Chr	position	Nearest gene	Coded/ Non-coded Allele	The Han Chinese CLHLS						The European IDEAL				The U.S. NECS			
					Males			Females			Males		Females		Males		Females	
					MAF (case/ control)	*P*	Odds ratio	MAF (case/ control)	*P*	Odds ratio	*P*	Effect direct	*P*	Effect direct	*P*	Odds ratio	*P*	Odds ratio
Female-specific longevity loci																		
rs60210535	14	46635410	LINC00871	G/A	0.043/0.047	0.4863	0.872	0.031/0.050	9.0E-05	0.580	NA	NA	NA	NA	0.692	0.954	4.6E-5	0.700
rs2622624	4	89069406	ABCG2	T/C	0.385/0.339	0.0780	1.157	0.372/0.320	6.8E-05	1.237	0.586	+	0.003	+	0.236	1.114	0.278	0.928
rs11107893	12	95491530	FGD6	A/G	0.141/0.150	0.5547	0.934	0.163/0.129	9.8E-05	1.328	NA	NA	NA	NA	0.052	1.310	0.035	1.234
Male-specific longevity loci																		
rs11188697	10	98050113	DNTT	G/C	0.097/0.052	2.0E-05	1.937	0.083/0.079	0.4437	1.075	0.038	+	0.172	+	0.293	1.174	0.942	0.992
rs618454	2	97080093	NCAPH	C/G	0.160/0.217	7.2E-05	0.6628	0.192/0.201	0.3577	0.9421	NA	NA	NA	NA	0.049	0.827	0.635	1.034
rs138863	22	50204273	BRD1	G/A	0.010/0.039	9.5E-06	0.2242	0.021/0.026	0.1688	0.792	0.454	-	0.587	+	0.045	0.723	0.438	0.920

Notes: (1) In the European IDEAL, “+” means the allele is more frequent in individuals ≥85 years of age as compared to individuals <65 years of age, while "-" means the opposite. (2) NA: not available; MAF: minor allele frequency.

SM Table 3. The 11 *male-specific* pathways significantly enriched and associated with longevity (P<0.005 and FDR<0.05)

Pathway/Gene set name	Description	*P*-value	FDR	Significant genes/ selected genes/All genes
CELLULAR BIOSYNTHETIC PROCESS	Genes annotated by the GO term GO:0044249. The chemical reactions and pathways resulting in the formation of substances, carried out by individual cells.	< 0.001	0.0	8/23/321
GTPASE REGULATOR ACTIVITY	Genes annotated by the GO term GO:0030695. Modulates the rate of GTP hydrolysis by a GTPase.	< 0.001	0.0040	4/16/126
INFLAMMATORY RESPONSE	Genes annotated by the GO term GO:0006954. The immediate defensive reaction (by vertebrate tissue) to infection or injury caused by chemical or physical agents. The process is characterized by local vasodilation, extravasation of plasma into intercellular spaces and accumulation of white blood cells and macrophages.	< 0.001	0.00525	5/14/130
TRANSLATION	Genes annotated by the GO term GO:0006412. The chemical reactions and pathways resulting in the formation of a protein. This is a ribosome-mediated process in which the information in messenger RNA (mRNA) is used to specify the sequence of amino acids in the protein.	< 0.001	0.005666667	4/14/180
BIOSYNTHETIC PROCESS	Genes annotated by the GO term GO:0009058. The energy-requiring part of metabolism in which simpler substances are transformed into more complex ones, as in growth and other biosynthetic processes.	< 0.001	0.0066666664	9/36/470
LIGASE ACTIVITY	Genes annotated by the GO term GO:0016874. Catalysis of the ligation of two substances with concomitant breaking of a diphosphate linkage, usually in a nucleoside triphosphate. Ligase is the systematic name for any enzyme of EC class 6.	< 0.001	0.0068000006	2/11/97
MACROMOLECULE BIOSYNTHETIC PROCESS	Genes annotated by the GO term GO:0009059. The chemical reactions and pathways resulting in the formation of macromolecules, large molecules including proteins, nucleic acids and carbohydrates.	< 0.001	0.009428571	6/27/321
ENZYME REGULATOR ACTIVITY	Genes annotated by the GO term GO:0030234. Modulates the activity of an enzyme.	0.0020	0.03522222	6/33/324
RNA BINDING	Genes annotated by the GO term GO:0003723. Interacting selectively with an RNA molecule or a portion thereof.	0.0010	0.038125	4/16/259
IMMUNE RESPONSE	Genes annotated by the GO term GO:0006955. Any immune system process that functions in the calibrated response of an organism to a potential internal or invasive threat.	< 0.001	0.041100003	4/17/237
SMOOTH MUSCLE CONTRACTION	This pathway illustrates signaling networks implicated in uterine muscle contraction at labor and quiescence throughout gestation (pregnancy).	0.0030	0.049909092	3/19/150

SM Table 4. The 34 *female-specific* pathways significantly enriched and associated with longevity (P<0.005 and FDR<0.05)

Pathway/Gene set name	Description	*P*-value	FDR	Significant genes/ Selected genes/All genes
TUBULIN BINDING	Genes annotated by the GO term GO:0015631. Interacting selectively with monomeric or multimeric forms of tubulin, including microtubules.	< 0.001	0.00925	4/14/47
TRYPTOPHAN METABOLISM	tryptophan => kynurenine => 2-aminomuconate	< 0.001	0.009666666	3/11/56
RESPONSE TO NUTRIENT LEVELS	Genes annotated by the GO term GO:0031667. A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus reflecting the presence, absence, or concentration of nutrients.	< 0.001	0.0108	4/11/29
VASCULATURE DEVELOPMENT	Genes annotated by the GO term GO:0001944. The process whose specific outcome is the progression of the vasculature over time, from its formation to the mature structure.	< 0.001	0.011333334	6/19/55
MICROTUBULE BINDING	Genes annotated by the GO term GO:0008017. Interacting selectively with microtubules, filaments composed of tubulin monomers.	0.0020	0.0117500005	2/10/33
CATION BINDING	Genes annotated by the GO term GO:0043169. Interacting selectively with cations, charged atoms or groups of atoms with a net positive charge.	< 0.001	0.0124285715	12/61/213
PGC1APATHWAY	PCG-1a is expressed in skeletal muscle, heart muscle, and brown fat, and is a coactivator for receptors such as glucocorticoid receptor and thyroid hormone receptor.	0.0010	0.012727273	3/11/24
COAGULATION	Genes annotated by the GO term GO:0050817. The process by which a fluid solution, or part of it, changes into a solid or semisolid mass.	< 0.001	0.012846153	5/14/44
HSA04920 ADIPOCYTOKINE SIGNALING PATHWAY	Genes involved in adipocytokine signaling pathway	< 0.001	0.013	5/24/72
ANGIOGENESIS	Genes annotated by the GO term GO:0001525. Blood vessel formation when new vessels emerge from the proliferation of pre-existing blood vessels.	< 0.001	0.0133125	5/17/48
REGULATION OF BODY FLUID LEVELS	Genes annotated by the GO term GO:0050878. Any process that modulates the levels of body fluids.	< 0.001	0.013333334	5/14/57
HEMOSTASIS	Genes annotated by the GO term GO:0007599. The stopping of bleeding (loss of body fluid) or the arrest of the circulation to an organ or part.	< 0.001	0.013333334	5/14/48
ZINC ION BINDING	Genes annotated by the GO term GO:0008270. Interacting selectively with zinc (Zn) ions.	< 0.001	0.0138	4/23/90
TRANSITION METAL ION BINDING	Genes annotated by the GO term GO:0046914. Interacting selectively with a transition metal ions; a transition metal is an element whose atom has an incomplete d-subshell of extranuclear electrons, or which gives rise to a cation or cations with an incomplete d-subshell. Transition metals often have more than one valency state. Biologically relevant transition metals include vanadium, manganese, iron, copper, cobalt, nickel, molybdenum and silver.	< 0.001	0.014	6/27/111
SMALL GTPASE REGULATOR ACTIVITY	Genes annotated by the GO term GO:0005083. Modulates the rate of GTP hydrolysis by a small monomeric GTPase.	< 0.001	0.016	5/17/67
MITOCHONDRIAL ENVELOPE	Genes annotated by the GO term GO:0005740. The double lipid bilayer enclosing the mitochondrion and separating its contents from the cell cytoplasm; includes the intermembrane space.	0.001	0.016529411	3/12/97
ORGAN MORPHOGENESIS	Genes annotated by the GO term GO:0009887. Morphogenesis of an organ. An organ is defined as a tissue or	< 0.001	0.017052632	11/50/145

	set of tissues that work together to perform a specific function or functions. Morphogenesis is the process by which anatomical structures are generated and organized. Organs are commonly observed as visibly distinct structures, but may also exist as loosely associated clusters of cells that work together to perform a specific function or functions.			
POSITIVE REGULATION OF DEVELOPMENTAL PROCESS	Genes annotated by the GO term GO:0051094. Any process that activates or increases the rate or extent of development, the biological process whose specific outcome is the progression of an organism over time from an initial condition (e.g. a zygote, or a young adult) to a later condition (e.g. a multicellular animal or an aged adult).	< 0.001	0.017222222	13/61/218
PROTEIN HETERODIMERIZATION ACTIVITY	Genes annotated by the GO term GO:0046982. Interacting selectively with a nonidentical protein to form a heterodimer.	< 0.001	0.017590908	7/27/77
PROTEIN DIMERIZATION ACTIVITY	Genes annotated by the GO term GO:0046983. The formation of a protein dimer, a macromolecular structure consists of two noncovalently associated identical or nonidentical subunits.	< 0.001	0.017695652	15/64/182
HSA00340 HISTIDINE METABOLISM	Genes involved in histidine metabolism	< 0.001	0.017904762	4/11/41
RESPONSE TO EXTRACELLULAR STIMULUS	Genes annotated by the GO term GO:0009991. A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of an extracellular stimulus.	< 0.001	0.018041667	4/12/33
GUANYL NUCLEOTIDE EXCHANGE FACTOR ACTIVITY	Genes annotated by the GO term GO:0005085. Stimulates the exchange of guanyl nucleotides by a GTPase. Under normal cellular physiological conditions, the concentration of GTP is higher than that of GDP, favoring the replacement of GDP by GTP in association with the GTPase.	< 0.001	0.01805	5/17/47
RIBONUCLEASE ACTIVITY	Genes annotated by the GO term GO:0004540. Catalysis of the hydrolysis of phosphodiester bonds in chains of RNA.	0.0020	0.01988	3/10/25
HSA00380 TRYPTOPHAN METABOLISM	Genes involved in tryptophan metabolism	< 0.001	0.022	6/20/60
ORGANELLE ENVELOPE	Genes annotated by the GO term GO:0031967. A double membrane structure enclosing an organelle, including two lipid bilayers and the region between them. In some cases, an organelle envelope may have more than two membranes.	< 0.001	0.022555556	7/28/169
ENVELOPE	Genes annotated by the GO term GO:0031975. A multilayered structure surrounding all or part of a cell; encompasses one or more lipid bilayers, and may include a cell wall layer, also includes the space between layers.	< 0.001	0.022555556	7/28/169
BLOOD COAGULATION	Genes annotated by the GO term GO:0007596. The sequential process by which the multiple coagulation factors of the blood interact, ultimately resulting in the formation of an insoluble fibrin clot; it may be divided into three stages: stage 1, the formation of intrinsic and extrinsic prothrombin converting principle; stage 2, the formation of thrombin; stage 3, the formation of stable fibrin polymers.	0.0010	0.031535715	4/13/43
ION BINDING	Genes annotated by the GO term GO:0043167. Interacting selectively with ions, charged atoms or groups of atoms.	< 0.001	0.0334	13/74/274
NEGATIVE REGULATION OF MULTICELLULAR ORGANISMAL PROCESS	Genes annotated by the GO term GO:0051241. Any process that stops, prevents or reduces the frequency, rate or extent of an organismal process, the processes pertinent to the function of an organism above the cellular level; includes the integrated processes of tissues and organs.	0.0040	0.033724137	2/11/32
REGULATION OF APOPTOSIS	Genes annotated by the GO term GO:0042981. Any process that modulates the occurrence or rate of cell death by apoptosis.	0.0010	0.03951613	18/88/342
ANATOMICAL STRUCTURE	Genes annotated by the GO term GO:0048646. The process pertaining to the initial formation of an anatomical structure	0.0020	0.04063636	5/22/56

FORMATION	from unspecified parts. This process begins with the specific processes that contribute to the appearance of the discrete structure and ends when the structural rudiment is recognizable. An anatomical structure is any biological entity that occupies space and is distinguished from its surroundings. Anatomical structures can be macroscopic such as a carpel, or microscopic such as an acrosome.			
REGULATION OF PROGRAMMED CELL DEATH	Genes annotated by the GO term GO:0043067. Any process that modulates the frequency, rate or extent of programmed cell death, cell death resulting from activation of endogenous cellular processes.	0.0010	0.040705882	18/89/343
REGULATION OF DEVELOPMENTAL PROCESS	Genes annotated by the GO term GO:0050793. Any process that modulates the frequency, rate or extent of development, the biological process whose specific outcome is the progression of a multicellular organism over time from an initial condition (e.g. a zygote, or a young adult) to a later condition (e.g. a multicellular animal or an aged adult).	< 0.001	0.04165625	25/132/442

SM Table 5. PRS analysis of the 11 *male-specific* longevity top loci (replicated & $P<10^{-5}$ in males, but P>0.05 in females)

	(A) PRS analysis of the 11 male-specific longevity top loci in **males**				(B) PRS analysis of the 11 male-specific longevity top loci in **females**				(C) PRS analysis of the 11 male-specific longevity top loci in **sexes-mixed** dataset, adjusted for sex			
Trisection of The PRS	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P
High PRS	0.0647	0.0252	2.91	1.6E-14	0.0004	0.0017	1.11	0.220	0.0167	0.0075	1.38	1.4E-5
Middle PRS	0.0152	0.0121	1	--	-0.0029	0.0008	1	--	0.0027	0.003	1	--
Low PRS	-0.0355	0.0271	0.37	1.1E-10	-0.0074	0.0032	0.9	0.229	-0.0123	0.0089	0.69	1.7E-7
Pseudo R^2	0.112				0.001				0.015			
	Percent distribution of PRS scores				Percent distribution of PRS scores				Percent distribution of PRS scores			
	centenarians	Mid-age controls			centenarians	Mid-age controls			centenarians	Mid-age controls		
High PRS	52.7	19.7			35.1	31.5			39.5	28.1		
Middle PRS	31.6	34.3			33.3	33.4			33.4	32.7		
Low PRS	15.7	46			31.6	35.1			27.1	39.2		
# participants	564	773			1614	1526			2178	2299		

Notes: (1) The 11 male-specific longevity top loci are listed in Table 1. (2) As explained in the SM section S1, the odds ratios in this and all other Tables and Figures cannot be interpreted as the size of pure effects of the genotype on longevity.

SM Table 6. PRS analysis of the 12 *female-specific* longevity top loci (replicated & $P<10^{-5}$ in females, but P>0.05 in males)

	(A) PRS analysis of the 12 female-specific longevity top loci in **females**				(B) PRS analysis of the 12 female-specific longevity top loci in **males**				(C) PRS analysis of the 12 female-specific longevity top loci in **sexes-mixed** dataset, adjusted for sex			
Trisection of The PRS	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P
High PRS	0.0316	0.0141	2.07	9.8E-16	0.0011	0.0023	1.14	0.344	0.0198	0.0091	1.66	8.4E-12
Middle PRS	0.0046	0.0058	1.00	--	-0.0044	0.0015	1.00	--	0.0018	0.0039	1.00	--
Low PRS	-0.0232	0.0152	0.56	1.0E-10	-0.0106	0.0035	0.91	0.470	-0.0173	0.0101	0.66	1.3E-08
Pseudo R^2	0.049				0.002				0.025			
	Percent distribution of PRS scores				Percent distribution of PRS scores				Percent distribution of PRS scores			
	centenarians	Mid-age controls			centenarians	Mid-age controls			centenarians	Mid-age controls		
High PRS	43.9	22.2			35.6	31.7			41.4	25.7		
Middle PRS	32.6	34.1			33.2	33.5			32.8	33.8		
Low PRS	23.5	43.7			31.2	34.8			25.8	40.5		
# participants	1614	1526			564	773			2178	2299		

Notes: (1) The 12 female-specific longevity top loci are listed in Table 2. (2) The same as those in Table 1.

SM Table 7. PRS analysis of the 44 *male-specific* longevity strong loci ($10^{-5} \leq P < 10^{-4}$ in males but $P>0.4$ in females)

	(A) PRS analysis of the 44 male-specific longevity strong loci in **males**				(B) PRS analysis of the 44 male-specific longevity strong loci in **females**				(C) PRS analysis of the 44 male-specific longevity strong loci in **sexes- mixed dataset**, adjused for sex			
Trisection of The PRS	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P
High PRS	0.0452	0.0129	8.33	2.8E-41	-0.0002	0.0006	1.14	0.1367	0.01	0.0031	1.89	3.6E-17
Middle PRS	0.0182	0.006	1.00	--	-0.0015	0.0003	1.00	--	0.0039	0.0013	1.00	--
Low PRS	-0.0063	0.0109	0.21	6.4E-17	-0.0028	0.0007	0.90	0.2293	-0.0017	0.0027	0.77	0.0012
Pseudo R^2	0.290				0.002				0.0289			
	Percent distribution of PRS scores				Percent distribution of PRS scores				Percent distribution of PRS scores			
	centenarians	Mid-age controls			centenarians	Mid-age controls			centenarians	Mid-age controls		
High PRS	64.5	10.6			35.3	31.3			41.4	25.8		
Middle PRS	27.5	37.7			33.2	33.4			31.5	35.0		
Low PRS	8.0	51.7			31.5	35.3			27.1	39.2		
# participants	564	773			1614	1526			2178	2299		

SM Table 8. PRS analysis of the 58 *female-specific* longevity strong loci ($10^{-5} \leq P < 10^{-4}$ in females but $P > 0.35$ in males)

	(A) PRS analysis of the 58 female-specific longevity strong loci in **females**				(B) PRS analysis of the 58 female-specific longevity strong loci in **males**				(C) PRS analysis of the 58 female-specific longevity strong loci in **sexes- mixed dataset**, adjustied for sex			
Trisection of The PRS	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P
High PRS	0.014	0.0065	4.58	3.9E-51	0.001	0.0011	1.27	0.0801	0.0084	0.0045	2.94	8.4E-43
Middle PRS	0.0015	0.0029	1.00	--	-0.0009	0.0004	1.00	--	-0.0001	0.0019	1.00	--
Low PRS	-0.0111	0.0062	0.30	2.9E-35	-0.0027	0.0009	0.83	0.1793	-0.0087	0.0042	0.45	4.0E-24
Pseudo R^2	0.179				0.005				0.100			
	Percent distribution of PRS scores				Percent distribution of PRS scores				Percent distribution of PRS scores			
	centenarians	Mid-age controls			centenarians	Mid-age controls			centenarians	Mid-age controls		
High PRS	53.0	12.5			37.6	30.3			49.2	18.3		
Middle PRS	32.1	34.7			33.0	33.6			31.5	35.1		
Low PRS	14.9	52.8			29.4	36.1			19.3	46.6		
# participants	1614	1526			564	773			2178	2299		

SM Table 9. PRS analysis of the 191 *male-specific* longevity moderate loci ($10^{-4} \leq P < 10^{-3}$ in males but $P > 0.75$ in females)

	(A) PRS analysis of the 191 male-specific longevity moderate loci in **males**				(B) PRS analysis of the 191 male-specific longevity moderate loci in **females**				(C) PRS analysis of the 191 male-specific longevity moderate loci in **sexes- mixed dataset**, adjusted for sex			
Trisection of The PRS	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P
High PRS	0.0131	0.0058	82.75	6.1E-49	0.0007	0.0001	1.15	0.1153	0.003	0.0015	2.50	1.1E-31
Middle PRS	-0.0028	0.0039	1.00	--	0.0004	0.0001	1.00	--	0	0.0006	1.00	--
Low PRS	-0.0165	0.0051	0.02	4.0E-12	0.0001	0.0001	0.96	0.6152	-0.0028	0.0014	0.57	2.3E-13
Pseudo R^2	0.622				0.001				0.067			
	Percent distribution of PRS scores				Percent distribution of PRS scores				Percent distribution of PRS scores			
	centenarians	Mid-age controls			centenarians	Mid-age controls			centenarians	Mid-age controls		
High PRS	76.8	1.7			35.1	31.5			46.0	21.4		
Middle PRS	22.7	41.1			32.8	33.9			32.2	34.3		
Low PRS	0.5	57.2			32.1	34.6			21.8	44.3		
# participants	564	773			1614	1526			2178	2299		

SM Table 10. PRS analysis of the 311 female-specific longevity moderate loci ($10^{-4} \leq P < 10^{-3}$ in females but $P > 0.7$ in males)

	(A) PRS analysis of the 311 female-specific longevity moderate loci in **females**				(B) PRS analysis of the 311 female-specific longevity moderate loci in **males**				(C) PRS analysis of the 311 female-specific longevity moderate loci in **sexes-mixed dataset**, adjusted for sex			
Trisection of The PRS	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P	PRS mean	PRS SD	OR	P
High PRS	0.0062	0.0026	16.64	2.1E-73	-0.0002	0.0002	1.29	0.0595	0.0037	0.0018	5.00	7.7E-74
Middle PRS	0.0000	0.0017	1.00	--	-0.0006	0.0001	1.00	--	-0.0004	0.001	1.00	--
Low PRS	-0.0065	0.0026	0.05	3.3E-89	-0.0009	0.0002	0.84	0.2022	-0.0044	0.0018	0.18	1.0E-76
Pseudo R^2	0.468				0.006				0.248			
	Percent distribution of PRS scores				Percent distribution of PRS scores				Percent distribution of PRS scores			
	centenarians	Mid-age controls			centenarians	Mid-age controls			centenarians	Mid-age controls		
High PRS	61.6	3.4			37.8	30.1			56.9	11.0		
Middle PRS	34.7	31.9			32.8	33.8			32.7	34.0		
Low PRS	3.7	64.7			29.4	36.1			10.4	55.0		
# participants	1614	1526			564	773			2178	2299		

SM Table 11. All independent loci associated with longevity at a significance level of $P<10^{-5}$, $P<10^{-4}$ or $P<10^{-3}$ in males or females, and compared to the results in the other sex

	32 loci associated with longevity In males or females ($P<10^{-5}$), and compared to the other sex								207 loci associated with longevity in males or females ($P<10^{-4}$), and compared to the other sex								1,665 loci associated with longevity In males or females ($P<10^{-3}$), and compared to the other sex							
	13 loci in **males** $P<10^{-5}$		Results of these 13 loci in females		19 loci in **females** $P<10^{-5}$		Results of these 19 loci in males		88 loci in **males** $P<10^{-4}$		Results of these 88 loci in females		119 loci in **females** $P<10^{-4}$		Results of these 119 loci in males		733 loci in **males** $P<10^{-3}$		Results of these 733 loci in females		933 loci in **females** $P<10^{-3}$		Results of these 933 loci in males	
P-value	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
$P<10^{-6}$	2	15.4	0	0	0	0	0	0	2	2.3	0	0	0	0	0	0	2	0.3	0	0	0	0	0	0
$10^{-6}\le P<10^{-5}$	11	84.6	0	0	19	100	0	0	11	12.5	0	0	19	16.0	0	0	11	1.5	0	0	19	2.0	1	0.1
$10^{-5}\le P<10^{-4}$	0	0	0	0	0	0	0	0	75	85.2	0	0	100	84.0	0	0	75	10.2	0	0	100	10.7	0	0
$10^{-4}\le P<10^{-3}$	0	0	1	7.7	0	0	0	0	0	0	1	1.1	0	0	0	0	645	88.0	1	0.1	814	87.3	0	0
$10^{-3}\le P<0.05$	0	0	0	0	0	0	5	26.3	0	0	6	6.8	0	0	10	8.4	0	0	52	7.1	0	0	55	5.9
P>=0.05	0	0	12	92.3	0	0	14	73.7	0	0	81	92.1	0	0	109	91.6	0	0	680	92.8	0	0	877	94.0
Total	13	100	13	100	19	100	19	100	88	100	88	100	119	100	119	100	733	100	733	100	933	100	933	100

SM Table 12. Percentage distributions of PRS scores of High-PRS, Mid-PRS and Low-PRS of each of the groups of the loci associated with longevity among male and female centenarians and middle-aged controls, analyzed in Figs 1-3 and SM Figs 5-7

Sex-specific longevity loci	Males						Females					
	High-PRS		Mid-PRS		Low-PRS		High-PRS		Mid-PRS		Low-PRS	
	Cent.	Contr.	Cent.	Contr.	Cent.	Contr.	Cent.	Contr.	Cent.	Contr.	Cent.	Contr.
Sex-specific loci associated with longevity												
11 male-specific longevity top loci ($P<10^{-5}$ in males but P>0.05 in females) analyzed in Fig 1(a)	54.26	21.09	30.67	30.40	15.07	48.51	33.71	31.39	34.88	34.14	31.41	34.47
44 male-specific longevity top loci ($10^{-5}\le P<10^{-4}$ in males but $P>0.4$ in females) analyzed in Fig 2(a)	72.52	14.88	19.68	33.38	7.8	51.75	30.48	31.19	35.63	35.91	33.89	32.9
191 male-specific longevity strong loci ($10^{-4}\le P<10^{-3}$ in males but $P>0.75$ in females) analyzed in Fig 3(a)	87.23	4.4	11.88	24.84	0.89	70.76	31.54	29.95	39.34	39.19	29.12	30.87
12 female-specific longevity strong loci ($P<10^{-5}$ in females but P>0.05 in males) analyzed in Fig 1(b)	32.80	31.31	34.57	35.71	32.62	32.99	44.42	22.80	32.22	32.83	23.36	44.36
58 female-specific longevity moderate loci ($10^{-5}\le P<10^{-4}$ in females but $P>0.35$ in males) analyzed in Fig 2(b)	33.69	27.81	35.99	39.33	30.32	32.86	54.58	13.5	29.86	32.96	15.55	53.54
311 female-specific longevity moderate loci ($10^{-4}\le P<10^{-3}$ in females but $P>0.7$ in males) analyzed in Fig 3(b)	29.26	24.32	43.79	45.92	26.95	29.75	66.48	4.33	28.87	27.79	4.65	67.89
All loci associated with longevity												
32 loci associated with longevity ($P<10^{-5}$) in males or females analyzed in SM Fig 5	43.62	22.9	35.82	30.92	20.57	46.18	44.73	22.74	33.71	33.22	21.56	44.04
207 loci associated with longevity ($P<10^{-4}$) in males or females analyzed in SM Fig 6	57.98	13.32	31.74	36.61	10.28	50.06	56.63	9.7	32.4	33.22	10.97	57.08
1,665 loci associated with longevity ($P<10^{-3}$) in males or females analyzed in SM Fig 7	92.55	12.94	NA	NA	7.45	87.06	94.30	6.23	NA	NA	5.70	93.77

Note: Cent.: centenarians; Contr.: middle-aged controls.

SM Table 13. Estimators of the sex-specific genetic & relative benefit/loss ratios

Sex	OR_{GE}: odds ratio of longevity for those with a combination of G (genotype) and E (sex)			Formulas of the sex-specific genetic & relative benefit/loss ratios
	G=1: Low-PRS	G=2: Mid-PRS	G=3: High-PRS	
E=1: Males (M)	OR_{11}	OR_{21}=1.0	OR_{31}	
E=2: Females (F)	OR_{12}	OR_{22}	OR_{32}	
From females' perspective:(F-M)/M	*F1 =abs. value of* $(OR_{12}-OR_{11})/OR_{11}$	*F2 =abs. value of* $(OR_{22}-OR_{21})/OR_{21}$	*F3 =abs. value of* $(OR_{32}-OR_{31})/OR_{31}$	Female genetic & relative benefit/loss ratio: (F3 / F1) for own-sex-specific longevity loci (F1 / F3) for other-sex longevity loci
From males' perspective:(M-F)/F	*M1= abs. value of* $(OR_{11}-OR_{12})/OR_{12}$	*M2= abs. value of* $(OR_{21}-OR_{22})/OR_{22}$	*M3= abs. value of* $(OR_{31}-OR_{32})/OR_{32}$	Male genetic & relative benefit/loss ratio: (M3 / M1) for own-sex-specific longevity loci ((M1 / M3) for other-sex longevity loci

SM Table 14a. Power estimates for the *male-specific GWAS* based on the CLHLS data: the minimum detectable odds ratios of the effects size of genotypes to reach a statistical power >80%, using QUANTO (1.1) software to perform the power calculation

Ideal p-value	$p<5\times10^{-8}$	$p<5\times10^{-7}$	$p<5\times10^{-6}$	$p<5\times10^{-5}$	$p<5\times10^{-4}$	$p<5\times10^{-3}$
Number of SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs
Analysis	Association with longevity	Association with longevity	Association with longevity	Association with longevity	Association with longevity	Association with longevity
Sample size	564 centenarians & 773 controls	564 centenarians & 773 controls	564 centenarians & 773 controls	564 centenarians & 773 controls	564 centenarians & 773 controls	564 centenarians & 773 controls
Dep. var.	Reaching age 100+	Reaching age 100+	Reaching age 100+	Reaching age 100+	Reaching age 100+	Reaching age 100+
Prevalence	$2.33\text{x}10^{-6}$	$2.33\text{x}10^{-6}$	$2.33\text{x}10^{-6}$	$2.33\text{x}10^{-6}$	$2.33\text{x}10^{-6}$	$2.33\text{x}10^{-6}$
MAF=0.05	2.52	2.39	2.25	2.11	1.95	1.78
MAF=0.10	2.05	1.97	1.87	1.78	1.67	1.55
MAF=0.20	1.78	1.71	1.65	1.57	1.50	1.41
MAF=0.30	1.68	1.63	1.57	1.50	1.44	1.36
MAF=0.40	1.65	1.60	1.54	1.48	1.41	1.34

SM Table 14b. Power estimates for the *female-specific* GWAS based on the CLHLS data: the minimum detectable odds ratios of the effects size of genotypes to reach a statistical power >80%, using QUANTO (1.1) software to perform the power calculation

Ideal p-value	$p<5\times10^{-8}$	$p<5\times10^{-7}$	$p<5\times10^{-6}$	$p<5\times10^{-5}$	$p<5\times10^{-4}$	$p<5\times10^{-3}$
Number of SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs	0.82M geno-typed & 4.78M imputed SNPs
Analysis	Association with longevity	Association with longevity	Association with longevity	Association with longevity	Association with longevity	Association with longevity
Sample size	1,614 centenarian & 1,526 controls	1,614 centenarian & 1,526 controls	1,614 centenarian & 1,526 controls	1,614 centenarian & 1,526 controls	1,614 centenarian & 1,526 controls	1,614 centenarian & 1,526 controls
Dep. var.	Reaching age 100+	Reaching age 100+	Reaching age 100+	Reaching age 100+	Reaching age 100+	Reaching age 100+
Prevalence	$7.83\text{x}10^{-6}$	$7.83\text{x}10^{-6}$	$7.83\text{x}10^{-6}$	$7.83\text{x}10^{-6}$	$7.83\text{x}10^{-6}$	$7.83\text{x}10^{-6}$
MAF=0.05	1.89	1.82	1.75	1.67	1.58	1.48
MAF=0.10	1.63	1.58	1.53	1.47	1.41	1.34
MAF=0.20	1.46	1.43	1.39	1.35	1.31	1.26
MAF=0.30	1.41	1.38	1.34	1.31	1.27	1.22
MAF=0.40	1.38	1.35	1.32	1.29	1.25	1.21

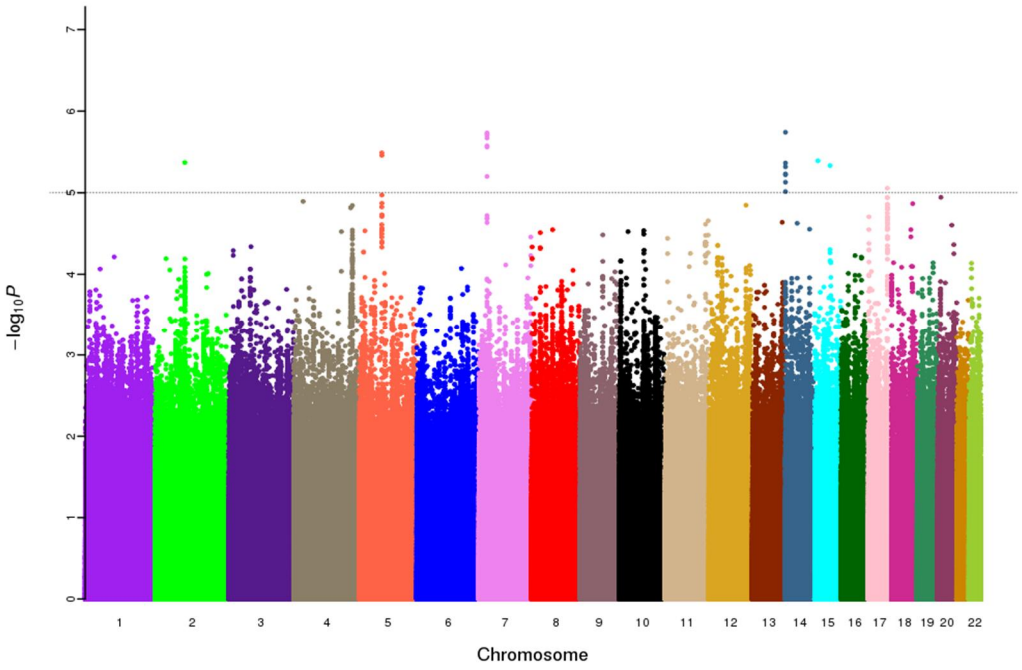
SM Table 15. Parameters used and outcome of power estimates for the sex-specific PRS analyses on longevity based on CLHLS GWAS data, using the *AVENGEME* software (27)

	Males			Females		
	11 loci ($P<10^{-5}$)	51 loci ($P<10^{-4}$)	136 loci ($P<10^{-3}$)	12 loci ($P<10^{-5}$)	65 loci ($P<10^{-4}$)	290 loci ($P<10^{-3}$)
nsnp	394,403	394,403	394,403	394,403	394,403	394,403
discovery/ target	794/543	794/543	794/543	1733/1407	1733/1407	1733/1407
vg1	0.15	0.15	0.15	0.15	0.15	0.15
cov12	0.03610	0.07558	0.38726	0.03650	0.07808	0.38726
$\pi0$	0.99999	0.99991	0.99910	0.99999	0.99993	0.99896
pupper	0, $1\text{x}10^{-5}$	0, $1\text{x}10^{-4}$	0, $1\text{x}10^{-3}$	0, $1\text{x}10^{-5}$	0, $1\text{x}10^{-4}$	0, $1\text{x}10^{-3}$
prevalence	$2.33\text{x}10^{-6}$	$2.33\text{x}10^{-6}$	$2.33\text{x}10^{-6}$	$7.83\text{x}10^{-6}$	$7.83\text{x}10^{-6}$	$7.83\text{x}10^{-6}$
sampling	0.3602	0.3602	0.3602	0.4784	0.4784	0.4784
alpha	0.05	0.05	0.05	0.05	0.05	0.05
nested	True	True	True	True	True	True
weighted	True	True	True	True	True	True
binary	True	True	True	True	True	True
lambdaS	--	--	--	--	--	--
shrinkage	False	False	False	False	False	False
logrisk	False	False	False	False	False	False
Power	**0.99747**	**0.99999**	**0.99915**	**1.00**	**1.00**	**1.00**

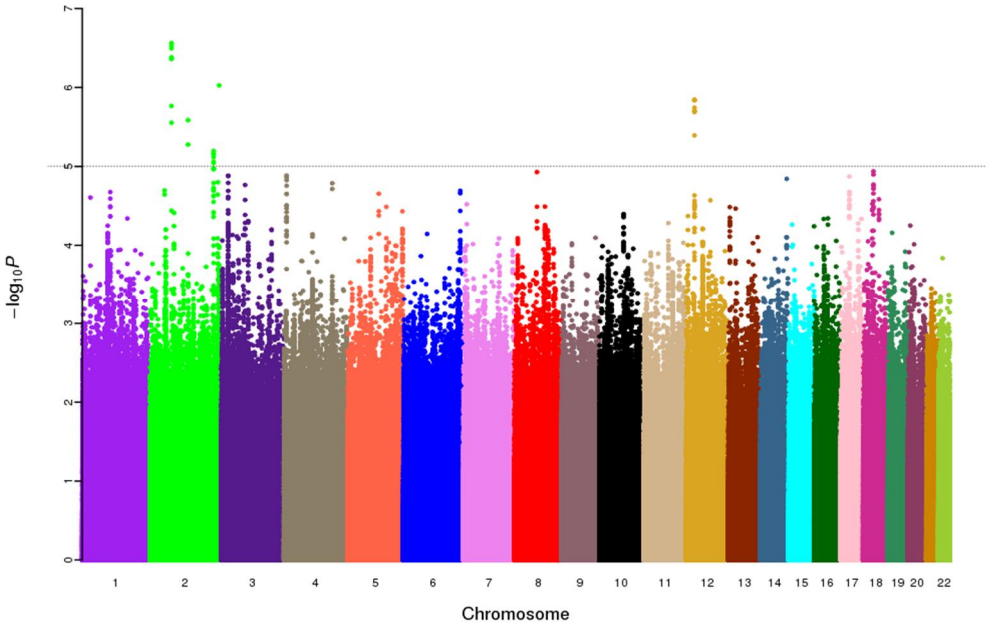
Notes: The following definitions of the parameters were taken from the Users' Manual of the ***AVENGEME*** software (27)

1) nsnp: number of independent SNPs in the GWAS. Clumping was performed from the total 5,595,657 SNPs using PLINK to retain SNPs with R2<0.1 within 250 kb windows.
2) discovery/target: size of discovery sample / size of target sample.
3) vg1: proportion of trait variance explained by the entire set of SNPs in the discovery sample.
4) cov12: covariance between genetic effects in the discovery and target samples, estimated by the estimatePolygenic Model function in AVENGEME.

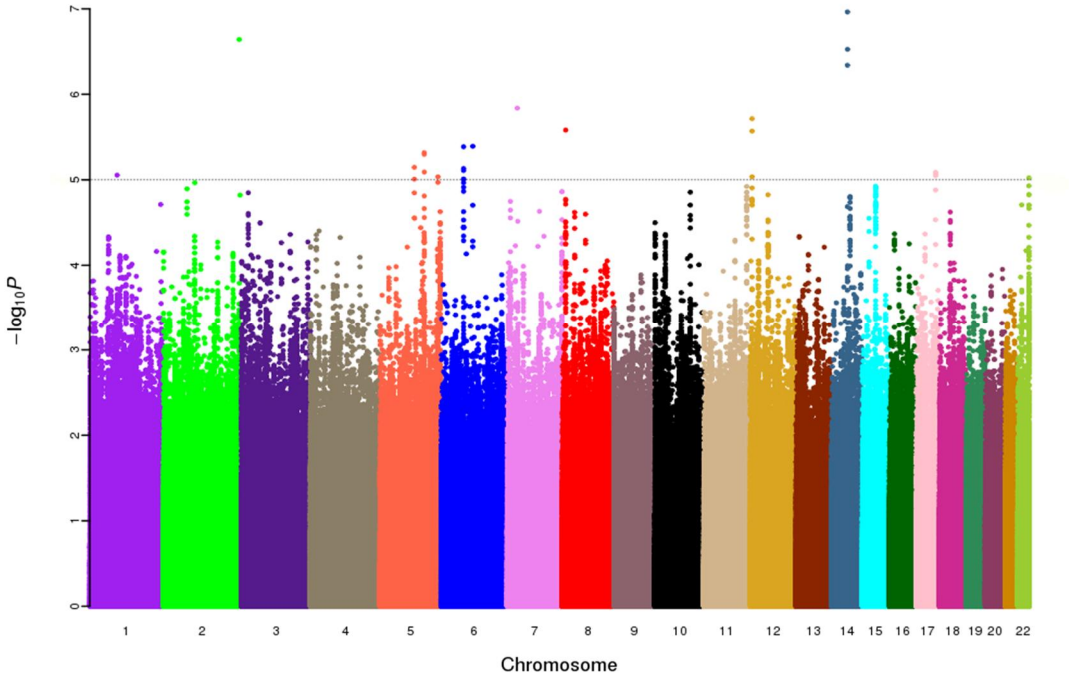
5) $\pi 0$ = 1- (# of sex-specific SNPs ($P<10^{-5}$, 10^{-4}, or 10^{-3} in one sex but P>0.05 in the other sex) among entire set of 5.6 million SNPs) / entire set of 5.6 million SNPs
6) pupper: lower bound, upper bound.
7) prevalence: Sex-specific prevalence rate of centenarians among the Chinese population.
8) sampling: (# of cases) / (# of total samples) in the discovery sample.
9) alpha: Default 0.05; no need to do Bonferroni correction since all selected SNPs are included in PRS and only one trait is considered (29).
10) nested: Only one interval.
11) weighted: PRS is constructed with weight *log(OR).*
12) binary: The trait "longevity" is binary.
13) shrinkage: Ridge regression model is not used.
14) logrisk: Liability scale is assumed.
15) The more detailed definitions of the parameters used for the power estimates of sex-specific PRS analyses are referred to the manual of the R program for PRS power estimates *AVENGEME* (27, 28)



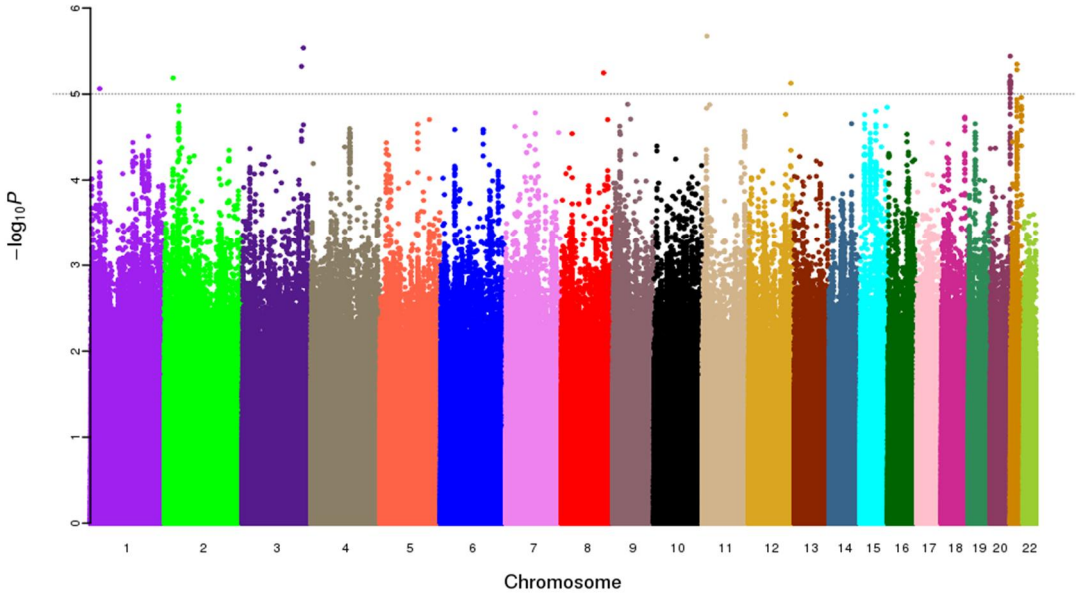
SM Fig 1a. Manhattan plot showing the results of the association with longevity in the *male discovery* (North region) GWAS dataset



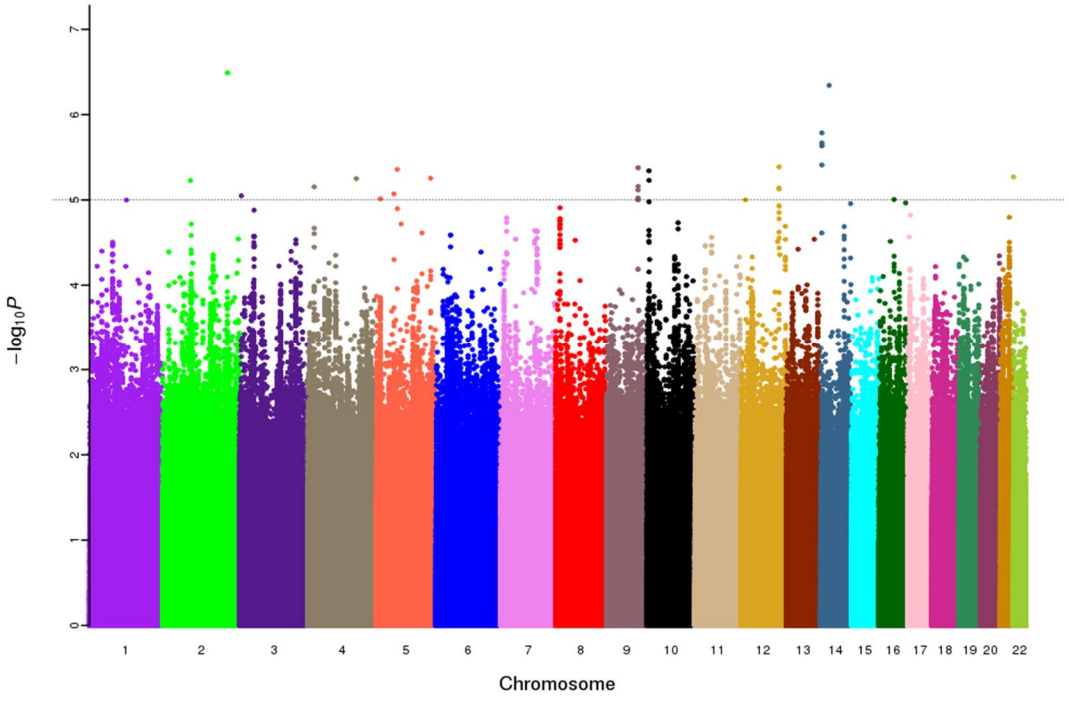
SM Fig 1b. Manhattan plot showing the results of the association with longevity in the *male evaluation* (South region) GWAS dataset



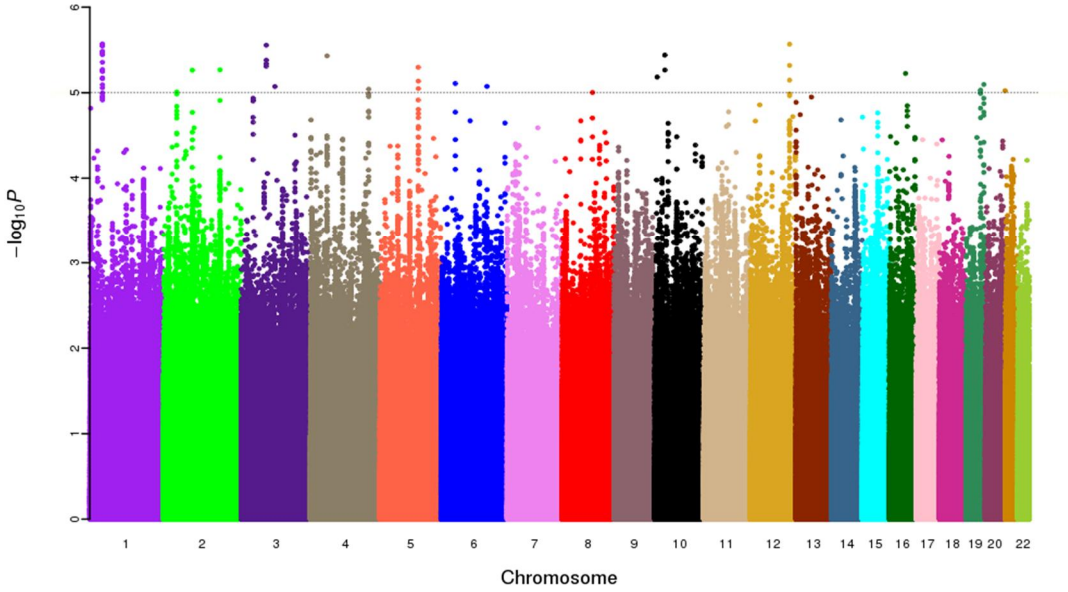
SM Fig 1c. Manhattan plot showing the results of the association with longevity in the *male North-South combined* GWAS dataset



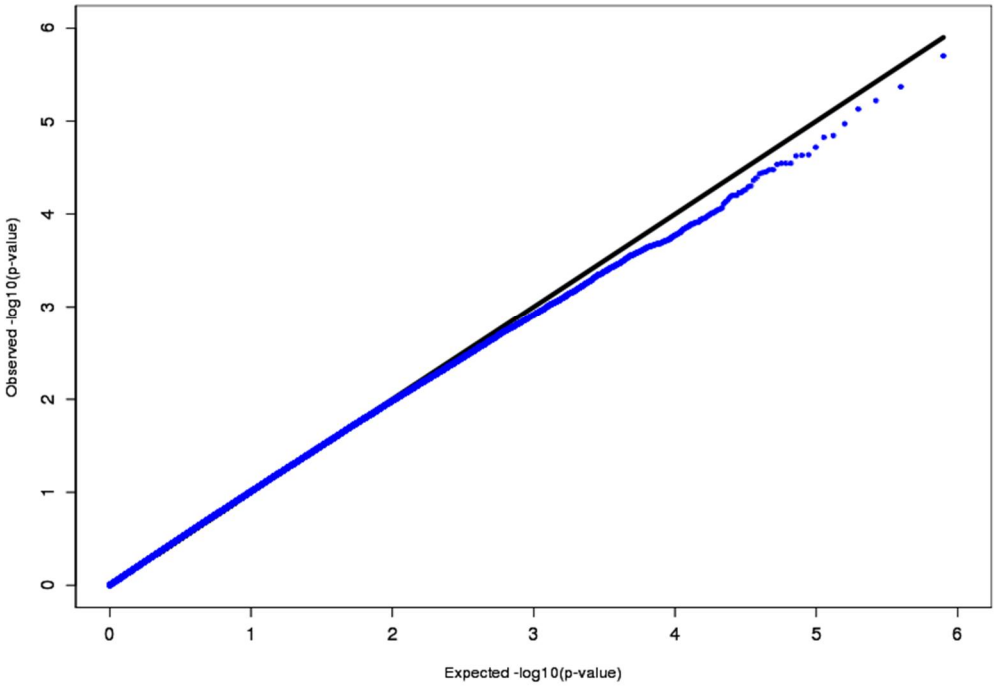
SM Fig 2a. Manhattan plot showing the results of the association with longevity in the *female discovery* (North region) GWAS dataset



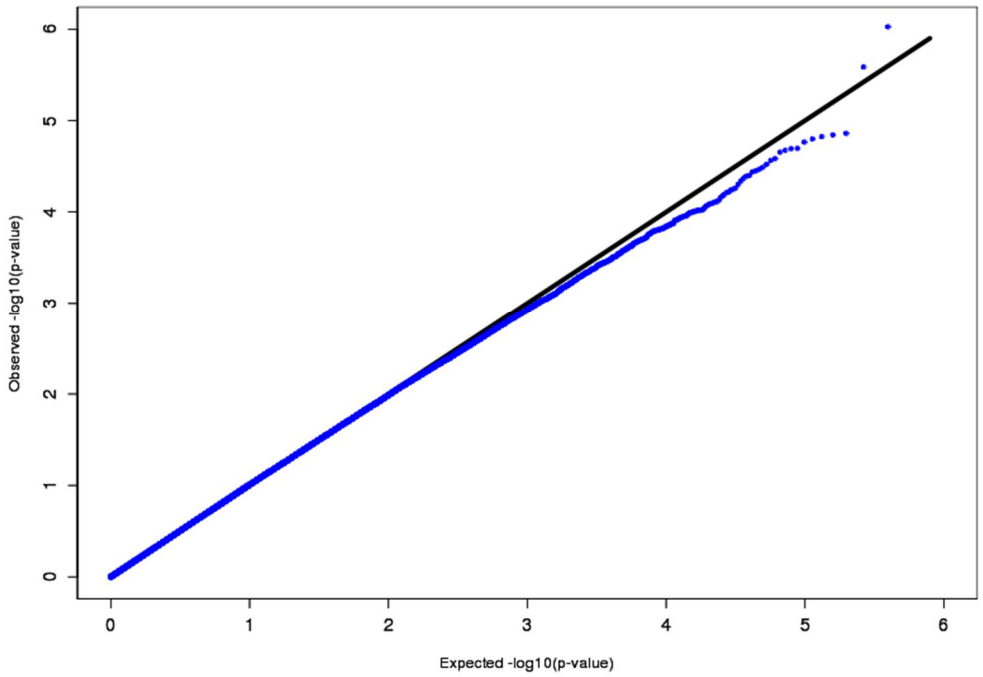
SM Fig 2b. Manhattan plot showing the results of the association with longevity in the *female evaluation* (South region) GWAS dataset



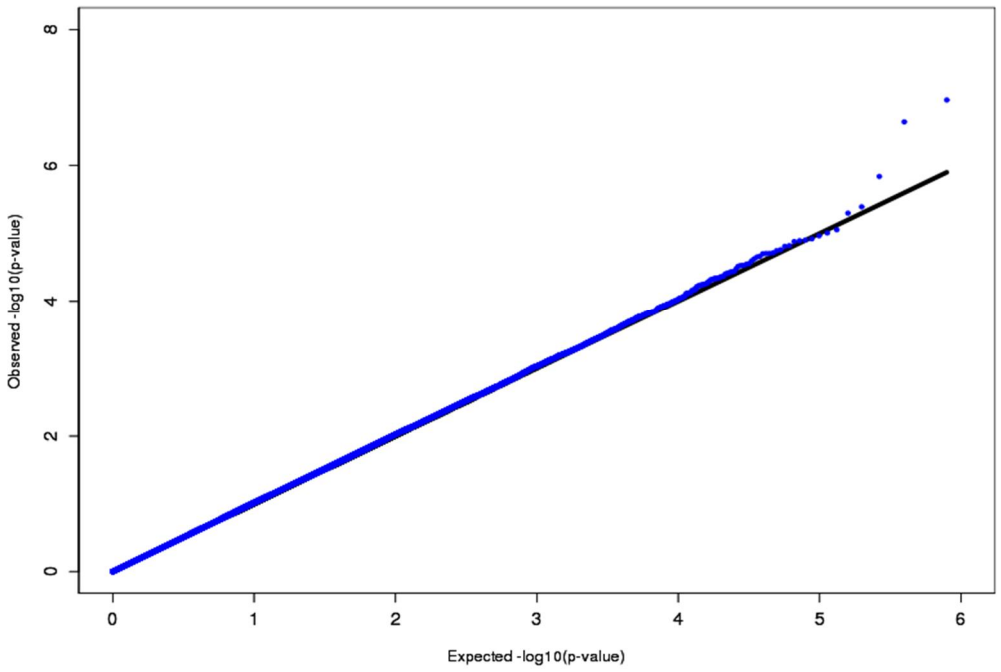
SM Fig 2c. Manhattan plot showing the results of the association with longevity in the *female North-South combined* GWAS dataset



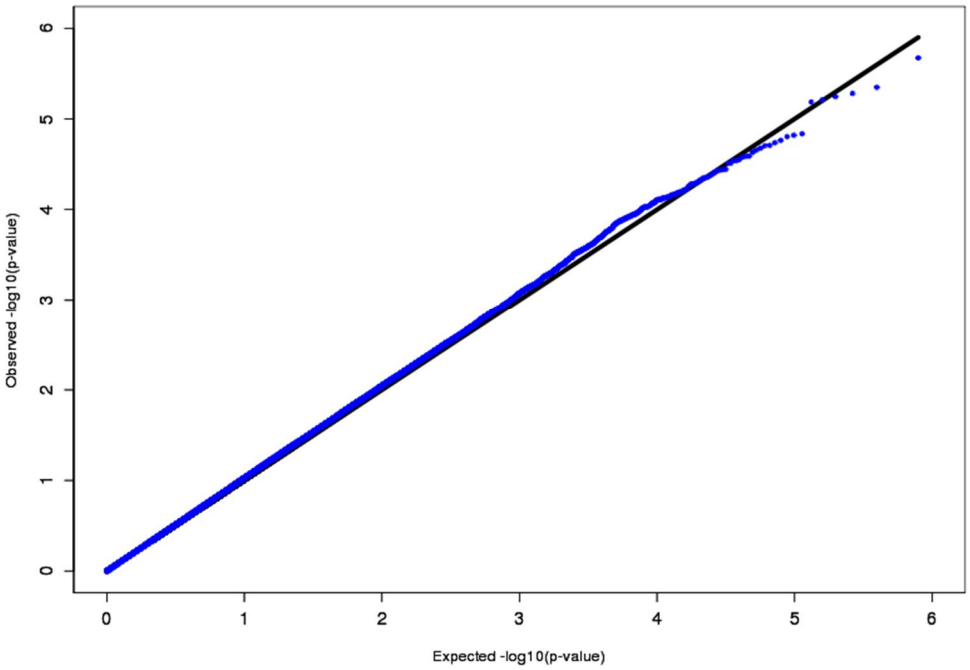
SM Fig 3a. Quantile-quantile plot in the *male discovery* (North region) GWAS dataset (The genomic inflation factor is 1.020 (λ=1.020))



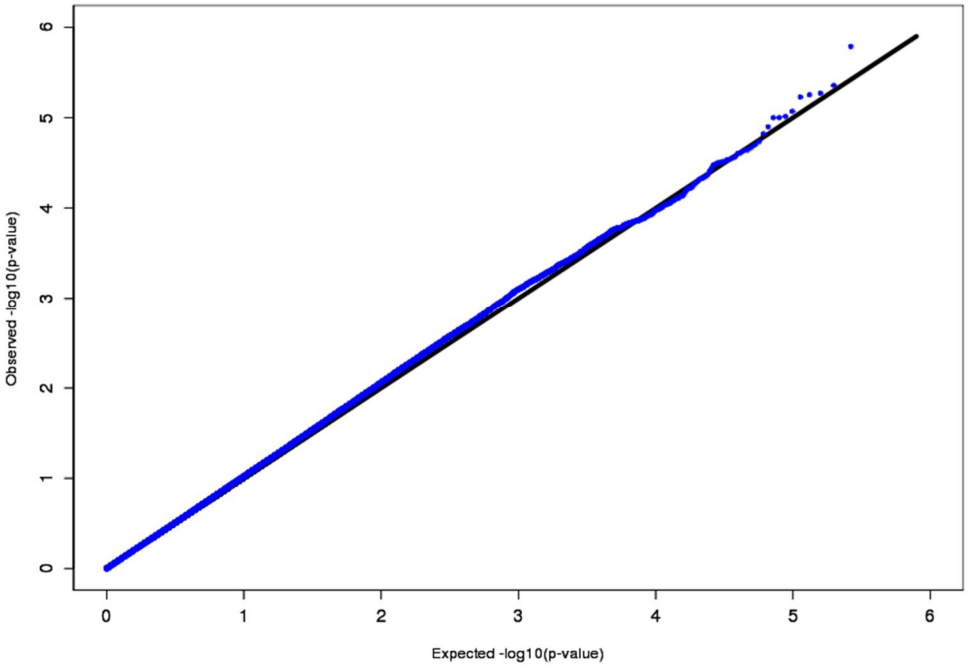
SM Fig 3b. Quantile-quantile plot in the *male evaluation* (South region) GWAS dataset (The genomic inflation factor is 1.010 (λ=1.010))



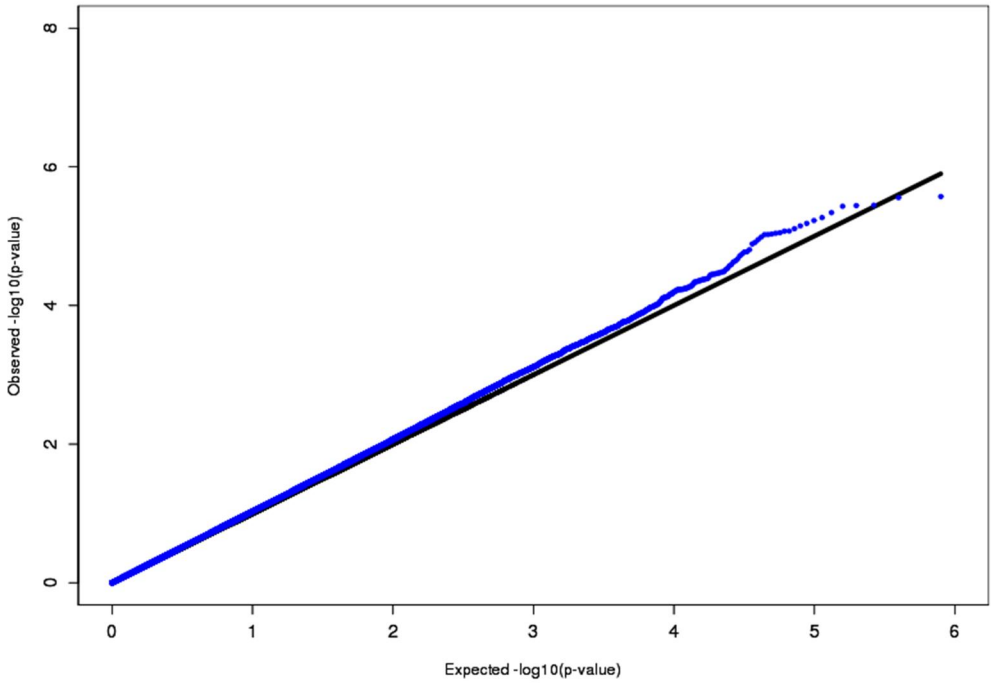
SM Fig 3c. Quantile-quantile plot in the *male North-South combined* GWAS dataset (The genomic inflation factor is 1.014 (λ=1.014))



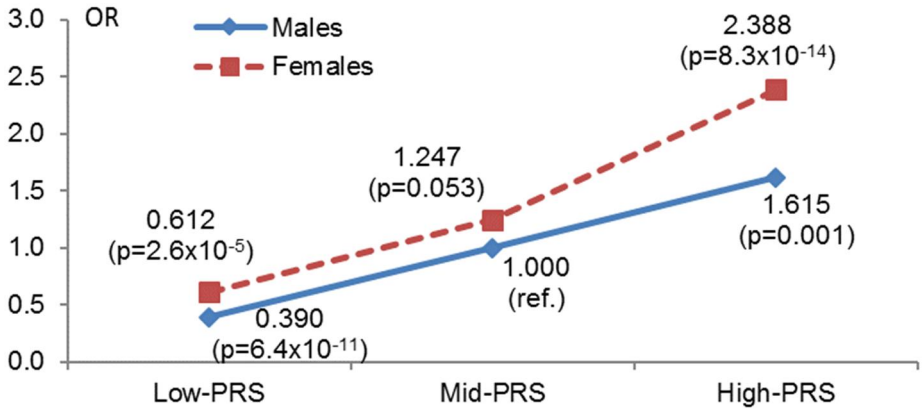
SM Fig 4a. Quantile-quantile plot in the *female discovery* (North region) GWAS dataset (The genomic inflation factor is 1.025 (λ=1.025)).



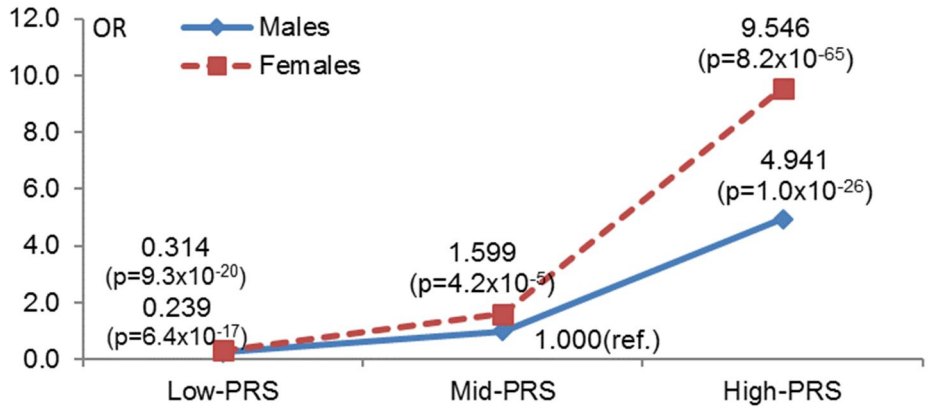
SM Fig 4b. Quantile-quantile plot in the *female evaluation* (South region) GWAS dataset (The genomic inflation factor is 1.016 (λ=1.016)).



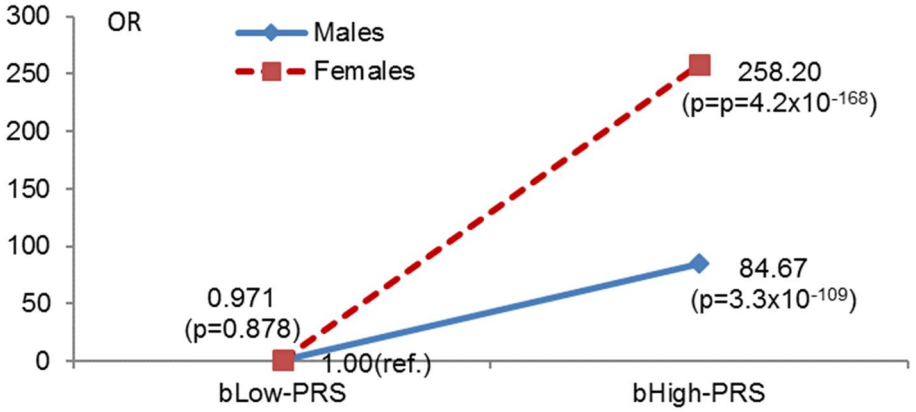
SM Fig 4c. Quantile-quantile plot in the *female North-South combined* GWAS dataset (The genomic inflation factor is 1.022 (λ=1.022)).



SM Fig 5. Odds ratios of longevity by the combinations of sex and trisections of the PRS summarizing 32 independent loci associated with longevity ($P<10^{-5}$) in males and/or females (including 13 loci in males and 19 loci in females)

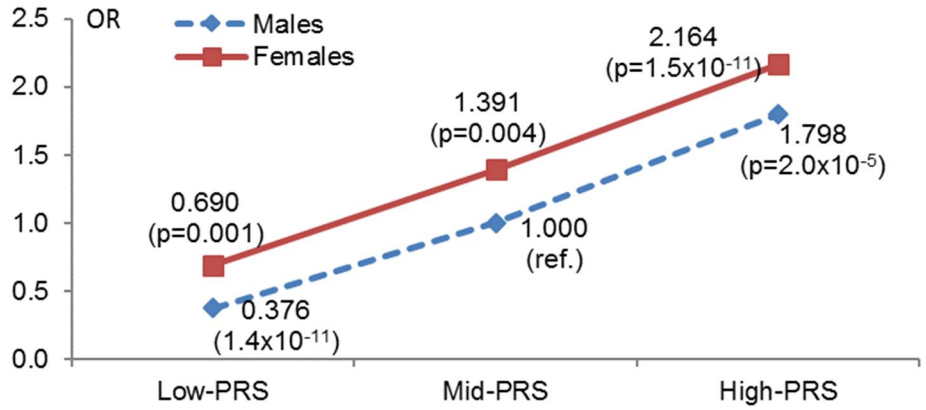


SM Fig 6. Odds ratios of longevity by the combinations of sex and trisections of the PRS summarizing 207 independent loci associated with longevity ($P<10^{-4}$) in males and/or females (including 88 loci in males and 119 loci in females)

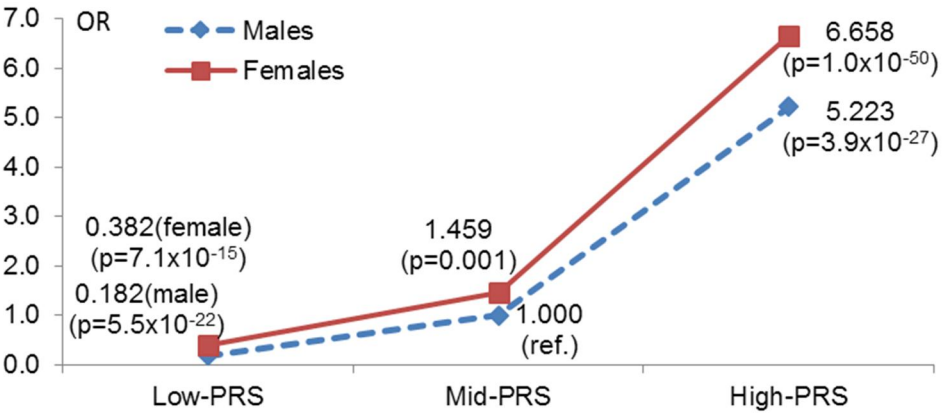


SM Fig 7. Odds ratios of longevity by the combinations of sex and binary variable of the PRS summarizing 1,665 independent loci associated with longevity (P<10-*3*) in males and/or females (including 733 loci in males and 933 loci in females)

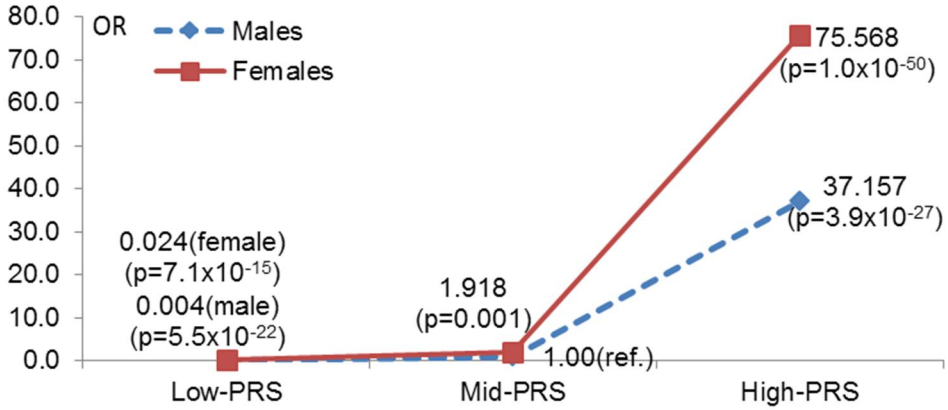
Note: If we used trisections of PRS for the 1,665 loci associated with longevity ($P<10^{-3}$) in males or females, there were only 5 male and 6 female centenarians with Low-PRS score, which implies that the sub-sample size of cases with Low-PRS is too small to produce statistically meaningful estimates. Thus, we use binary variable of PRS (with 50% cutoff) for the 1,665 loci with $P<10^{-3}$.



SM Fig 8. Odds ratios of longevity by the combinations of sex and trisections of the PRS summarizing 26 independent loci associated with longevity ($P<10^{-5}$) in males and/or females (including 13 male-specific loci and 13 randomly selected loci from the 19 female-specific loci)



SM Fig 9. Odds ratios of longevity by the combinations of sex and trisections of the PRS summarizing 176 independent loci associated with longevity ($P<10^{-4}$) in males and/or females (including 88 male-specific loci and 88 randomly selected loci from the 19 female-specific loci)



SM Fig 10. Odds ratios of longevity by the combinations of sex and binary variable of the PRS summarizing 1,466 independent loci associated with longevity (P<10-*3*) in males and/or females (including 733 male-specific loci and 733 randomly selected loci from the 932 female-specific loci)

References

1. Li Y, Wang WJ, Cao H, Lu J, Wu C, Hu FY, Guo J, Zhao L, Yang F, Zhang YX, Li W, Zheng GY, Cui H, Chen X, Zhu Z, He H, Dong B, Mo X, Zeng Y, Tian XL (2009) Genetic association of *FOXO1A* and *FOXO3A* with longevity trait in Han Chinese populations. *Hum Mol Genet* 18(24): 4897-4904.
2. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, Arnold AM, Aspelund T, Aulchenko YS, Benjamin EJ, Christiansen L, D'Agostino RB Sr, Fitzpatrick AL, Franceschini N, Glazer NL, Gudnason V, Hofman A, Kaplan R, Karasik D, Kelly-Hayes M, Kiel DP, Launer LJ, Marciante KD, Massaro JM, Miljkovic I, Nalls MA, Hernandez D, Psaty BM, Rivadeneira F, Rotter J, Seshadri S, Smith AV, Taylor KD, Tiemeier H, Uh HW, Uitterlinden AG, Vaupel JW, Walston J, Westendorp RG, Harris TB, Lumley T, van Duijn CM, Murabito JM (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.
3. Sebastiani P, Solovieff N, Dewan AT, Walsh KM, Puca A, Hartley SW, Melista E, Andersen S, Dworkis DA, Wilk JB, Myers RH, Steinberg MH, Montano M, Baldwin CT, Hoh J, Perls TT (2012) Genetic signatures of exceptional longevity in humans. *PLoS One* 7(1):e29848.
4. Zhao L, Yang F, Xu K, Cao H, Zheng GY, Zhang Y, Li J, Cui H, Chen X, Zhu Z, He H, Mo X, Kennedy BK, Suh Y, Zeng Y, Tian XL (2012) Common genetic variants of the β2-adrenergic receptor effect its translational efficiency and are associated with human longevity. *Aging Cell* 11(6):1094-1101.
5. Deelen, J. Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, Kamatani Y, Bennet AM, Tamm R, Trompet S, Guðbjartsson DF, Flachsbart F, Rose G, Viktorin A, Fischer K, Nygaard M, Cordell HJ, Crocco P, van den Akker EB, Böhringer S, Helmer Q, Nelson CP, Saunders GI, Alver M, Andersen-Ranberg K, Breen ME, van der Breggen R, Caliebe A, Capri M, Cevenini E, Collerton JC, Dato S, Davies K, Ford I, Gampe J, Garagnani P, de Geus EJ, Harrow J, van Heemst D, Heijmans BT, Heinsen FA, Hottenga JJ, Hofman A, Jeune B, Jonsson PV, Lathrop M, Lechner D, Martin-Ruiz C, Mcnerlan SE, Mihailov E, Montesanto A, Mooijaart SP, Murphy A, Nohr EA, Paternoster L, Postmus I, Rivadeneira F, Ross OA, Salvioli S, Sattar N, Schreiber S, Stefánsson H, Stott DJ, Tiemeier H, Uitterlinden AG, Westendorp RG, Willemsen G, Samani NJ, Galan P, Sørensen TI, Boomsma DI, Jukema JW, Rea IM, Passarino G, de Craen AJ, Christensen K, Nebel A, Stefánsson K, Metspalu A, Magnusson P, Blanché H, Christiansen L, Kirkwood TB, van Duijn CM, Franceschi C, Houwing-Duistermaat JJ, Slagboom PE (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.
6. Broer L, Buchman AS, Deelen J, Evans DS, Faul JD, Lunetta KL, Sebastiani P, Smith JA, Smith AV, Tanaka T, Yu L, Arnold AM, Aspelund T, Benjamin EJ, De Jager PL, Eirkisdottir G, Evans DA, Garcia ME, Hofman A, Kaplan RC, Kardia SL, Kiel DP, Oostra BA, Orwoll ES, Parimi N, Psaty BM, Rivadeneira F, Rotter JI, Seshadri S, Singleton A, Tiemeier H, Uitterlinden AG, Zhao W, Bandinelli S, Bennett DA, Ferrucci L, Gudnason V, Harris TB, Karasik D, Launer LJ, Perls TT, Slagboom PE, Tranah GJ, Weir DR, Newman

AB, van Duijn CM, Murabito JM (2015) GWAS of longevity in CHARGE consortium confirms *APOE* and *FOXO3* candidacy. *J Gerontol A Biol Sci Med Sci* 70(1):110–118.
7. Zeng Y, Nie C, Min J, Liu X, Li M, Chen H, Xu H, Wang M, Ni T, Li Y, Yan H, Zhang JP, Song C, Chi LQ, Wang HM, Dong J, Zheng GY, Lin L, Qian F, Qi Y, Liu X, Cao H, Wang Y, Zhang L, Li Z, Zhou Y, Wang Y, Lu J, Li J, Qi M, Bolund L, Yashin A, Land KC, Gregory S, Yang Z, Gottschalk W, Tao W, Wang J, Wang J, Xu X, Bae H, Nygaard M, Christiansen L, Christensen K, Franceschi C, Lutz MW, Gu J, Tan Q, Perls T, Sebastiani P, Deelen J, Slagboom E, Hauser E, Xu H, Tian XL, Yang H, Vaupel JW (2016) Novel loci and pathways significantly associated with longevity. *Sci Rep* 6, 21243. http://doi.org/10.1038/srep21243.
8. Zeng Y, Cheng L, Chen H, Cao H, Hauser ER, Liu Y, Xiao Z, Tan Q, Tian XL, Vaupel JW (2010) Effects of FOXO genotypes on longevity: a biodemographic analysis. *J Gerontol A Biol Sci Med Sci* 65(12): 1285-1299.
9. Zeng Y (2012) Towards Deeper Research and Better Policy for Healthy Aging--Using the Unique Data of Chinese Longitudinal Healthy Longevity Survey. *China Economic Journal* 5: 131-149.
10. Goodkind D (2009) Review on the book Healthy Longevity in China: Demographic, Socioeconomic, and Psychological Dimensions. *Population Studies* 63: 1-7.
11. Zeng Y, Cheng L, Zhao L, Tan Q, Feng Q, Chen H, Shen K, Li J, Zhang F, Cao H, Gregory SG, Yang Z, Gu J, Tao W, Tian XL, Hauser ER (2013) Interactions between Social/behavioral factors and ADRB2 genotypes may be associated with health at advanced ages in China. *BMC Geriatr* 13: 91-91.
12. Zeng Y, Chen H, Ni T, Ruan R, Feng L, Nie C, Cheng L, Li Y, Tao W, Gu J, Land KC, Yashin A, Tan Q, Yang Z, Bolund L, Yang H, Hauser E, Willcox DC, Willcox BJ, Tian XL, Vaupel JW (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.
13. Coale A.J, Li S (1991) The effect of age misreporting in China on the calculation of mortality rates at very high ages. *Demography* 28: 293-301.
14. Wang Z, Zeng Y, Jeune B, Vaupel JW (1998) Age validation of Han Chinese centenarians. *GENUS* 54: 123–141.
15. Zeng Y, Gu D (2008) Reliability of age reporting among the Chinese oldest-old in the CLHLS datasets. In: Zeng, Y. et al. (eds.). *Healthy longevity in China: Demographic, socioeconomic, and psychological dimensions.* pp.61-80. Dordrecht (The Netherlands: Springer Publisher).
16. Jia P, Wang L, Fanous AH, Pato CN, Edwards TL International Schizophrenia Consortium, Zhao Z (2012) Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia. *PLoS Comput Biol* 8(7):e1002587.
17. Institute of Medicine (US) Committee on Assessing Interactions Among Social, Behavioral, and Genetic Factors in Health; Hernandez LM, Blazer DG, editors. (2006) *Genes, Behavior, and the Social Environment: Moving Beyond the Nature/Nurture Debate*. (National Academies Press, Washington, D.C.)
18. Su MW, Tung KY, Liang PH, Tsai CH, Kuo NW, Lee YL (2012) Gene-gene and gene-environmental interactions of childhood asthma: a multifactor dimension reduction approach. *PLoS one* 7(2): e30694.
19. Ueki M (2014) On the choice of degrees of freedom for testing gene–gene interactions. *Stat Med* 33(28): 4934-4948.
20. Rudolph A, Milne RL, Truong T, Knight JA, Seibold P, Flesch-Janys D, Behrens S, Eilber U, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, Munday HR, Darabi H, Eriksson M, Brand JS, Olson J, Vachon CM, Hallberg E, Castelao JE, Carracedo A, Torres M, Li J, Humphreys K, Cordina-Duverger E, Menegaux F, Flyger H, Nordestgaard BG, Nielsen SF, Yesilyurt BT, Floris G, Leunen K, Engelhardt EG, Broeks A, Rutgers EJ, Glendon G, Mulligan AM, Cross S, Reed M, Gonzalez-Neira A, Arias Perez JI, Provenzano E, Apicella C, Southey MC, Spurdle A; kConFab Investigators; AOCS Group, Häberle L, Beckmann MW, Ekici AB, Dieffenbach AK, Arndt V, Stegmaier C, McLean C, Baglietto L, Chanock SJ, Lissowska J, Sherman ME, Brüning T, Hamann U, Ko YD, Orr N, Schoemaker M, Ashworth A, Kosma VM, Kataja V, Hartikainen JM, Mannermaa A, Swerdlow A; GENICA-Network, Giles GG, Brenner H, Fasching PA, Chenevix-Trench G, Hopper J, Benítez J, Cox A, Andrulis IL, Lambrechts D, Gago-Dominguez M, Couch F, Czene K, Bojesen SE, Easton DF, Schmidt MK, Guénel P, Hall P, Pharoah PD, Garcia-Closas M, Chang-Claude J (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.
21. Zeng Y, Chen H, Ni T, Ruan R, Nie C, Liu X, Feng L, Zhang F, Lu J, Li J, Li Y, Tao W, Gregory SG, Gottschalk W, Lutz MW, Land KC, Yashin A, Tan Q, Yang Z, Bolund L, Ming Q, Yang H,, Min J, Willcox DC, Willcox BJ, Gu J, Hauser E, Tian XL, Vaupel JW (2016) Interaction between FOXO1A-209 Genotype and Tea Drinking is Significantly Associated with Reduced Mortality at Advanced Ages. *Rejuvenation Res*

19(3):195-203.
22. Kendler KS, Kuhn JW, Vittum J, Prescott CA, Riley B (2005) The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: a replication. *Arch Gen Psychiatry* 62(5): 529-535.
23. Euesden J, Lewis CM, O'Reilly PF (2015) PRSice: polygenic risk score software. *Bioinformatics* 31(9): 1466-1468.
24. Levine ME. Crimmins EM, Prescott CA, Phillips D, Arpawong TE, Lee J (2014) A Polygenic Risk Score Associated with Measures of Depressive Symptoms Among Older Adults. *Biodemography Soc Biol* 60:199–211.
25. 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.
26. Population Census Office under the State Council and Department of Population, Social, Science and Technology Statistics, National Bureau of Statistics of the People's Republic of China (2002) *Tabulations of the 2000 Census of China*. (China Statistics Press, Beijing).
27. Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 9(3): e1003348.
28. Palla L, Dudbridge F (2015) A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am J Hum Genet* 97(2): 250-259.
29. Pouget JG, Han B, Mignot E, Ollila HM, Barker J, Spain S, Dand N, rembath R, Martin J, Mayes MD, Bossini-Castillo L, Lopez-Isac E, Jin Y, Santorico S, Spritz RA, Raychaudhuri S, Knight J (2016) Polygenic analysis of schizophrenia and 19 immune diseases reveals modest pleiotropy. http://dx.doi.org/10.1101/068684.