Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity

Marian Beekman[a,1], Christa Nederstigt[a], H. Eka D. Suchiman[a], Dennis Kremer[a], Ruud van der Breggen[a], Nico Lakenberg[a], Wendimagegn Ghidey Alemayehu[b], Anton J. M. de Craen[c], Rudi G. J. Westendorp[c], Dorret I. Boomsma[d], Eco J. C. de Geus[d], Jeanine J. Houwing-Duistermaat[b], Bastiaan T. Heijmans[a], and P. Eline Slagboom[a,e]

[a]Molecular Epidemiology, [b]Medical Statistics and Bioinformatics, and [c]Gerontology and Geriatrics, Leiden University Medical Center, 2300 RC, Leiden, The Netherlands; [d]Biological Psychology, Vrije Universiteit Amsterdam, 1081 BT, Amsterdam, The Netherlands; and [e]Netherlands Consortium for Healthy Ageing, 2300 RC, Leiden, The Netherlands

A set of currently known alleles increasing the risk for coronary artery disease, cancer, and type 2 diabetes as identified by genome-wide association studies was tested for compatibility with human longevity. Here, we show that nonagenarian siblings from long-lived families and singletons older than 85 y of age from the general population carry the same number of disease risk alleles as young controls. Longevity in this study population is not compromised by the cumulative effect of this set of risk alleles for common disease.

association | aging SNP

Members of long-lived families have a lifelong survival advantage (1, 2) that can be attributed to a lower risk for coronary artery disease, cancer, and type 2 diabetes (3–5). These observations may be explained by the presence of alleles protecting against diseases that contribute to population mortality or the absence of alleles promoting such diseases. The latter hypothesis can be tested now that genome-wide association studies (GWASs) have produced a series of SNPs showing robust associations with common diseases that are the main causes of death.

We tested whether nonagenarians from the Leiden Longevity Study (LLS) and octogenarians from the Leiden 85 Plus Study carried lower total numbers of 30 major GWAS-identified disease susceptibility alleles than younger controls.

Results

Reviewing the catalog of published GWASs for coronary artery disease, heart failure, cancer, and type 2 diabetes up to February 2009 revealed 30 SNPs in 22 disease-associated loci (Table S1). These SNPs were investigated for their deleterious role in survival beyond 85 y of age by comparing 723 nonagenarian siblings (mean age of 94 y) from the LLS (1), representing familial long-lived cases, and 721 unrelated younger controls (mean age of 52 y). We also compared 979 singleton long-lived individuals older than 85 y (mean age of 87 y) of age from the population-based Leiden 85 Plus Study (6, 7) and 1,167 younger controls (mean age of 41 y) from the Netherlands Twin Register (NTR) (8).

We tested whether the number of disease alleles was different between long-lived individuals and younger controls, because the cumulative effect of disease alleles is expected to be more pronounced than that of individual SNPs, with a modest effect on disease risk (9, 10). On average the nonagenarian siblings carried 26.7 ± 0.19 disease risk alleles, which was virtually identical to the number carried by younger controls (27.0 ± 0.12; $P = 0.127$; Fig. 1A), and the same was observed for the number carried by sporadic long-lived subjects (26.8 ± 0.11) and younger controls (26.8 ± 0.10; $P = 0.847$; Fig. 1B). The distribution of the number of risk alleles was the same in highly aged and young subjects, even in the tails of the distribution (Fig. 1 C and D). Similar results were obtained when the 19 alleles associated with meta-

bolic disease or the 11 cancer-associated alleles were considered separately. We compared 391 long-lived subjects and 409 younger controls in the 10% tails of the sum of risk allele distribution. The carriers of the 10% upper tail of the sum of risk allele distributions did not have a decreased risk of surviving beyond 85 y of age, because the odds ratio (OR) estimates are close to unity (Tables S2–S4). The 95% confidence interval (CI) of the OR per extra risk allele ranges from 0.97 to 1.01. Further exploration for different effect sizes and interacting alleles did not reveal any evidence for a subset of alleles associating with longevity via main effects or as interactions.

In a meta-analysis of the two cohorts, nominal significant associations with survival into old age were observed for seven individual SNPs ($0.023 < P < 0.046$; Table S5), but none was significant when accounting for multiple testing ($P < 0.001$).

The effect on fatal disease can also be analyzed by a cross-sectional comparison of healthy groups of different age categories (11–13). Because disease risk alleles contribute to morbidity and mortality, the frequencies of those risk alleles are expected to decrease in the healthy elderly individuals of the population at large. However, the frequency of deleterious genotypes might increase among long-lived individuals because their protective genotype allows disease-related genes to accumulate (12). To investigate the presence of such a buffering mechanism, we divided all subjects and controls ($n = 3,590$) into 10 age groups (Table S6) and determined the mean number of risk alleles and the allele frequencies of the 30 SNPs in each age group. We did not observe differences in the mean number of risk alleles (Fig. S1) or in SNP-specific allele frequencies (Dataset S1).

Discussion

A lower genetic predisposition for disease in long-lived individuals may explain why long-lived families have a delayed age of onset and lower mortality attributable to common age-related diseases (3–5). Therefore, our aim was to investigate whether the absence of currently known genetic susceptibility alleles is essential for survival into old age. Hence, we compared the cumulative effect of alleles that have convincingly been identified as risk factors for cardiovascular disease, type 2 diabetes, and
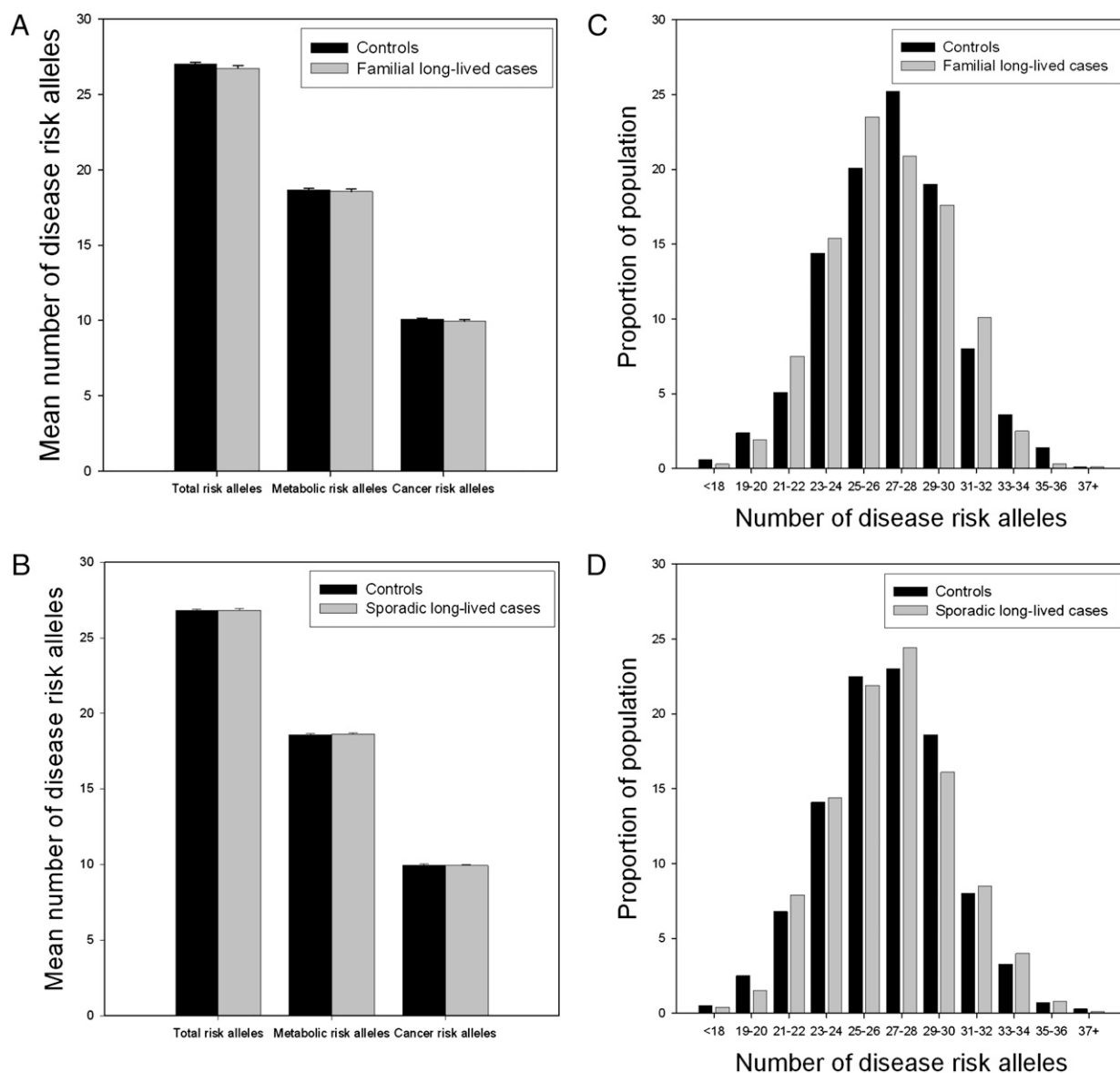
Fig. 1. GWAS-identified disease risk alleles in nonagenarians from long-lived families and sporadic long-lived individuals. (*A*) Average number of risk alleles among nonagenarian siblings and controls of the LLS. (*B*) Average number of risk alleles among sporadic long-lived individuals of the Leiden 85 Plus Study and NTR controls. (*C*) Distribution of the number of disease risk alleles among nonagenarian siblings and controls of the LLS. (*D*) Distribution of the number of disease risk alleles among sporadic long-lived participants of the Leiden 85 Plus Study and NTR controls.

cancer between long-lived individuals (aged 85 y and older) and the middle-aged general population, in whom the disease risk alleles are generally discovered.

Despite the fact that cardiovascular disease, cancer, and type 2 diabetes contribute to the majority of deaths in modern societies, the cumulative effects of this set of 30 risk alleles do not restrain the study group from surviving into old age. This observation, together with the observations that the offspring of the nonagenarian siblings and centenarians have a decreased prevalence of cardiovascular disease, hypertension, and type 2 diabetes as compared with controls (5, 14), suggests that the absence of these disease loci does not explain the lower morbidity in the offspring. Also, offspring of centenarians have lower mortality from cancer despite the same cancer prevalence as compared with controls (4). It may be possible that other factors, such as unidentified (rare) genetic variants that associate with fatal disease, are lacking in the study group. Alternatively, alleles protecting individuals from

cardiovascular disease, cancer, and type 2 diabetes by delaying the onset or decreasing the severity may be prevalent in this study group. To identify loci essential for delay of disease and survival, a GWAS for survival into old age may be an approach by which protective alleles could also be identified.

The sum of risk alleles has commonly been used to estimate a cumulative effect of the risk alleles to the disease prevalence of type 2 diabetes and cancer (9, 15–21). Previous studies reported substantially increased risk among the upper 10% of the population carrying the highest number of risk alleles as compared with the 10% with the lowest number of risk alleles [e.g., type 2 diabetes, OR= 8.7 (9); prostate cancer, OR = 9.5 (19)]. Of 20 known variants associated with type 2 diabetes, it has been estimated that they would account for 5–10% of an inherited predisposition for type 2 diabetes (22). The SNP sets that have been investigated for their cumulative effect on the risk for specific diseases in these studies overlap with the SNP set investigated in

the current study. No effect of a subset of SNPs or interactions among SNPs was detected in the current data, which are in agreement with previous observations that additive variance often accounts for almost all genetic variance (23). Putative associations are thus best described with a multiplicative model in line with previous disease-focused studies on GWAS SNP sets (9, 15–21). An accurate estimation of a predicted cumulative effect of these SNPs on longevity was hardly possible because the genetic associations used for this estimation derive mostly from studies focused on disease risk and not on the effect of disease-specific mortality.

In our study, we detected the odds of the carriers with the highest number of risk alleles, with an OR of 0.99 (95% CI: 0.97–1.03), to survive into old age. The boundaries of the 95% CI in this analysis are thus close to unity, which was also the case for the boundaries of the OR per extra risk allele. Although, in general, the association of disease risk alleles with longevity remains to be investigated (if all common and uncommon disease risk alleles are discovered and tested for their association with human longevity accounting for their interaction with environmental factors), our data suggest that longevity is not likely determined by the absence of GWAS-identified risk alleles for cardiovascular disease, cancer, and type 2 diabetes.

Because the current GWAS-identified SNPs have a relatively low predictive value for disease risk (24–26), it was expected that such risk alleles only marginally affect population-wide survival. However, the cumulative effect of cardiovascular disease, cancer, and type 2 diabetes risk alleles, which was expected to be more pronounced (9, 10), did not affect population survival. The extent to which disease risk attributable to the cumulative effects of susceptibility alleles is associated with population-based mortality risk should be tested in patient-based studies. Alternatively, other genetic and environmental factors present in individuals surviving into old age may counteract the detrimental effects of disease susceptibility alleles.

Materials and Methods

Study Populations. All participants of the LLS, Leiden 85 Plus Study, and NTR provided written informed consent after receiving an explanation of the nature and consequences of the study.

LLS and Control Samples. For the LLS, nonagenarian siblings of European descent were recruited together with their offspring and the partners of the offspring, who served as population controls. Families were recruited if at least two long-lived siblings were alive and fulfilled the age criterion of 89 y or older for men and 91 y or older for women, representing less than 0.5% of the Dutch population in 2001 (1). In total, 944 long-lived proband siblings with a mean age of 94 y (range: 89–104 y), 1,671 offspring with a mean age of 60 y (range: 39–81 y), and 744 partners with a mean age of 60 y (range: 36–79 y) were included.

Because we recently obtained Illumina Human660W-Quad and Human-OmniExpress GWAS data for participants in the LLS, we were able to investigate potential substructure. First, in PLINK version 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) (27) for 10,000 random SNPs, IBS (identity-by-state) estimates for all pairs of subjects in the samples were computed. Next, we performed a multidimensional scaling (mds) analysis on the N × N matrix of genome-wide IBS pairwise distances, using the mds-plot option in conjunction with clustering. The C1 values were plotted against C2 values to identify clustering among the subjects and controls (Fig. S2). We conclude that there is no substructure among the participants of the LLS to an extent that would affect our conclusions.

Leiden 85 Plus Study. In the Leiden 85 Plus Study, two prospective population-based cohorts consisting of inhabitants of Leiden aged 85 y and older were followed. Between 1987 and 1989 (cohort 1), 673 subjects aged 85 y and older of Dutch ancestry and white ethnicity were enrolled and followed up for survival for 17 y; during this time, 672 subjects (99.9%) died. Between 1997 and 1999, 563 subjects were enrolled in the month of their 85th birthday (cohort 2) and followed up for survival for 10 y; during this time, 453 subjects (81%) died. Subjects were visited at their home, and there were no exclusion criteria related to health. DNA was available from the combined cohorts for 1,245 subjects aged 85 y and older.

NTR as Population Controls. From the NTR, 1,203 unrelated participants of European descent with mean age of 41 y (range: 18–80 y) were selected as young population controls for whom DNA was available (8). The substructure in the NTR control group has already been reported (28), and in our study, we included the samples of Dutch descent without known family relations (i.e., those without any substructure).

SNP Selection. Using the catalog of published GWASs published up to February 2009 (http://www.genome.gov/26525384), we reviewed 266 GWASs, 69 of which reported on associations with coronary artery disease, heart failure, cancer, and type 2 diabetes. These 69 GWASs reported 23 disease-associated loci resulting from at least 2 independent GWASs, harboring 81 associated SNPs with $P < 10^{-4}$ (Table S1). For each locus, the most replicated SNP was selected from the 81 SNPs, and, subsequently, in case of an equal number of replications, the SNP with the lowest reported P value was selected. For 5 loci, 9 further disease-associated SNPs were selected that were in low to moderate linkage disequilibrium ($r^2 < 0.80$) with the most replicated SNP. Thus, 32 SNPs were selected to be genotyped with the Sequenom iPLEX. Because 2 SNPs failed in genotyping, 30 SNPs covering 22 loci were studied. Seventeen loci are represented by 1 SNP, 3 loci are represented by 2 SNPs, 1 locus is represented by 3 SNPs (8q24.21), and 1 locus is represented by 4 SNPs (9p21.3).

Genotyping. Genotyping of the selected SNPs was performed using the Sequenom MassARRAY iPLEXGold. The average genotype call rate for genotyped SNPs was 96.3%, and the average concordance rate was 99.7% among 4% duplicated control samples. All 30 SNPs were in Hardy–Weinberg equilibrium ($P \geq 0.002$) among the controls.

Statistical Analysis. *Single SNP analysis meta-analysis.* We applied a variance-modified Cochrane–Armitage test to take into account the relatedness between the highly aged sibling cases when computing the variances of the scores (29). In the meta-analysis of both the LLS and Leiden 85 Plus Study/NTR controls, the score statistics of the two studies were combined using a fixed-effects approach.
Cumulative SNP analysis. For each individual who was genotyped for all 30 SNPs, the total number of disease risk alleles was counted. The difference in the number of disease risk alleles between long-lived subjects and population controls was assessed by linear regression. In these analyses, the number of risk alleles was the dependent variable and the case status was included in the model as a categorical variable. To take into account dependencies within the long-lived sibships of the LLS, robust SEs were used [i.e., the variance was computed from the between-family variation (30)]. P values were also based on these robust SEs.

Logistical regression was performed with control/long-lived status as the outcome, study and gender as covariates, and 10% upper or lower tail as a variable. For the estimation of the OR per extra risk allele, logistical regression was performed with control/long-lived status as the outcome, study and gender as covariates, and the sum of risk alleles as a variable. Analyses were performed using the software package STATA/SE 11.0 (DPC Software).
Interaction analysis. Polychotomous regression (31) and logic regression (32) were applied to the datasets of the LLS and Leiden 85 Plus/NTR controls to identify subsets of SNPs that might have a joint effect on longevity. Cross-validation was used for selecting the best subset of SNPs and the best functional forms of those subsets. The two approaches are different in that the polychotomous regression is flexible in terms of inheritance modes of the SNPs, although only pairwise interactions are considered, whereas the logic regression is used to identify multiple interacting dominant and recessive effects of SNPs.

1. Schoenmaker M, et al. (2006) Evidence of genetic enrichment for exceptional survival using a family approach: The Leiden Longevity Study. *Eur J Hum Genet* 14:79–84.
2. Perls TT, et al. (2002) Life-long sustained mortality advantage of siblings of centenarians. *Proc Natl Acad Sci USA* 99:8442–8447.
3. Terry DF, Wilcox M, McCormick MA, Lawler E, Perls TT (2003) Cardiovascular advantages among the offspring of centenarians. *J Gerontol A Biol Sci Med Sci* 58:M425–M431.
4. Terry DF, et al. (2004) Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. *J Am Geriatr Soc* 52:2074–2076.
5. Westendorp RG, et al.; Leiden Longevity Study Group (2009) Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *J Am Geriatr Soc* 57:1634–1637.
6. Bootsma-van der Wiel A, et al. (2001) Disability in the oldest old: "Can do" or "do do"? *J Am Geriatr Soc* 49:909–914.
7. Kuningas M, et al. (2007) Haplotypes in the human Foxo1a and Foxo3a genes; Impact on disease and mortality at old age. *Eur J Hum Genet* 15:294–301.
8. Boomsma DI, et al. (2008) Genome-wide association of major depression: Description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet* 16:335–342.
9. Cauchi S, et al. (2008) Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value. *PLoS ONE* 3:e2031.
10. Lango H, et al.; UK Type 2 Diabetes Genetics Consortium (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes* 57:3129–3135.
11. Heijmans BT, Westendorp RG, Slagboom PE (2000) Common gene variants, mortality and extreme longevity in humans. *Exp Gerontol* 35:865–877.
12. Bergman A, Atzmon G, Ye K, MacCarthy T, Barzilai N (2007) Buffering mechanisms in aging: A systems approach toward uncovering the genetic component of aging. *PLOS Comput Biol* 3:e170.
13. Kammerer S, et al. (2003) Amino acid variant in the kinase binding domain of dual-specific A kinase-anchoring protein 2: A disease susceptibility polymorphism. *Proc Natl Acad Sci USA* 100:4066–4071.
14. Terry DF, Wilcox MA, McCormick MA, Perls TT (2004) Cardiovascular disease delay in centenarian offspring. *J Gerontol A Biol Sci Med Sci* 59:385–389.
15. Pharoah PD, Antoniou AC, Easton DF, Ponder BA (2008) Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* 358:2796–2803.
16. Gail MH (2008) Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *J Natl Cancer Inst* 100:1037–1041.
17. Sun J, et al. (2008) Cumulative effect of five genetic variants on prostate cancer risk in multiple study populations. *Prostate* 68:1257–1262.
18. Zheng SL, et al. (2008) Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 358:910–919.
19. Yamada H, et al. (2009) Replication of prostate cancer risk loci in a Japanese case-control association study. *J Natl Cancer Inst* 101:1330–1336.
20. Wacholder S, et al. (2010) Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 362:986–993.
21. Paynter NP, et al. (2010) Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 303:631–637.
22. McCarthy MI, Zeggini E (2009) Genome-wide association studies in type 2 diabetes. *Curr Diab Rep* 9:164–171.
23. Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4:e1000008.
24. Goldstein DB (2009) Common genetic variation and human traits. *N Engl J Med* 360:1696–1698.
25. Hirschhorn JN (2009) Genomewide association studies—Illuminating biologic pathways. *N Engl J Med* 360:1699–1701.
26. Kraft P, Hunter DJ (2009) Genetic risk prediction—Are we there yet? *N Engl J Med* 360:1701–1703.
27. Purcell S, et al. (2007) PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81:559–575.
28. Sullivan PF, et al. (2009) Genome-wide association for major depressive disorder: A possible role for the presynaptic protein piccolo. *Mol Psychiatry* 14:359–375.
29. Slager SL, Schaid DJ (2001) Evaluation of candidate genes in case-control studies: A statistical method to account for related subjects. *Am J Hum Genet* 68:1457–1462.
30. Diggle PJ, Liang KY, Zegers SL (1994) *Analysis of Longitudinal Data* (Oxford Univ Press, Oxford).
31. Kooperberg C, Ruczinski I (2005) Identifying interacting SNPs using Monte Carlo logic regression. *Genet Epidemiol* 28:157–170.
32. Ruczinski I, Kooperberg C, LeBlanc M (2003) Logic regression. *J Comput Graph Statist* 12:475–511.

GENETICS