Table 3
Relation of four SNPs to ischemic stroke determined by the chi-square test in subject panel C.

Gene	SNP (dbSNP)	Ischemic stroke[a]	Controls[a]	P (genotype)	P (allele)
CELSR1	A → G (rs6007897)			0.0265	0.0273
	AA	67 (94.4)	1737 (98.1)		
	AG	4 (5.6)	33 (1.9)		
	GG	0 (0)	0 (0)		
CELSR1	A → G (rs4044210)			0.0413	0.0425
	AA	67 (94.4)	1734 (98.0)		
	AG	4 (5.6)	36 (2.0)		
	GG	0 (0)	0 (0)		
LLGL2	T → C (rs1671021)			0.2332	0.1118
	TT	56 (78.9)	1277 (71.8)		
	TC	15 (21.1)	454 (25.5)		
	CC	0 (0)	48 (2.7)		
RUVBL2	C → T (rs1062708)			0.9385	0.7732
	CC	21 (29.6)	512 (28.9)		
	CT	35 (49.3)	852 (48.1)		
	TT	15 (21.1)	406 (22.9)		

[a] Numbers in parentheses are percentages.

type distribution and allele frequency, respectively), rs9615362 of CELSR1 (P = 0.0058 and 0.0060, respectively), and rs753307 of RUVBL2 (P = 0.0277 and 0.0239, respectively) — were significantly (P < 0.05) associated with ischemic stroke in terms of both genotype distribution and allele frequency in this population. We then sequenced exons and exon–intron boundaries of linkage disequilibrium (LD) blocks [standardized linkage disequilibrium coefficient (r^2) ≥ 0.3] including these SNPs (Supplementary Figure 1) in 96 DNA samples from subjects with ischemic stroke. Data for LD blocks were obtained from the International HapMap Project database (http://www.hapmap.org/index.html.ja).

The SNP rs1671021 (T → C, Phe479Leu) is located in exon 12 of LLGL2. Given that there were no data for LD blocks containing this SNP (Supplementary Figure 1A), we sequenced exon 12 of LLGL2 and its intron boundaries. Although there was no SNP other than rs1671021 in exon 12, rs2305526 (A → T) and a C → T polymorphism (not detected in dbSNP) were observed in intron 13. These three SNPs were in LD (Supplementary Table 4). The haplotype frequencies for these SNPs are shown in Supplementary Table 5. Both rs1671021 (T → C, Phe479Leu) in exon 12 and rs2305526 (A → T) in intron 13 were tag SNPs.

The SNP rs9615362 (A → C) is located in intron 10 of CELSR1. Given that an LD block containing this SNP includes exons 10–20 (Supplementary Figure 1B), we sequenced these exons and the corresponding exon–intron boundaries of CELSR1. In addition to rs9615362, we detected rs4044210 (A → G, Ile2107Val) in exon 17, rs6007897 (A → G, Thr2268Ala) in exon 20, and rs6008788 (A → G) in intron 19 of CELSR1. These four SNPs were in LD (Supplementary Table 6). The haplotype frequencies for these SNPs are shown in Supplementary Table 7. Both rs6007897 (A → G, Thr2268Ala) in exon 20 and rs6008788 (A → G) in intron 19 were tag SNPs.

The SNP rs753307 (C → T) is located in intron 13 of RUVBL2. Given that an LD block containing this SNP includes exons 7–15 of RUVBL2 and exon 3 of LHB (Supplementary Figure 1C), we sequenced these exons and the corresponding exon–intron boundaries. In addition to rs753307, we detected rs1062708 (C → T, Leu205Leu) in exon 8 and rs2287760 (C → G) in intron 8 of RUVBL2. These three SNPs were in LD (Supplementary Table 8). The haplotype frequencies for these SNPs are shown in Supplementary Table 9. Both rs1062708 (C → T, Leu205Leu) in exon 8 and rs2287760 (C → G) in intron 8 were tag SNPs.

Table 4
Relation of four SNPs to ischemic stroke determined by the chi-square test and multivariable logistic regression analysis in all subjects.

Gene	SNP (dbSNP)	Genotype distribution		Chi-square test		Multivariable logistic regression analysis			
		Ischemic stroke[a]	Controls[a]	P (genotype)	P (allele)	P (dominant)	OR (95% CI)	P (recessive)	OR (95% CI)
CELSR1	A → G (rs6007897)			0.0001	0.0001	0.0006	1.85 (1.29–2.61)		
	AA	940 (95.0)	5147 (97.3)						
	AG	50 (5.1)	144 (2.7)						
	GG	0 (0)	0 (0)						
	Hardy–Weinberg P	0.4150	0.3156						
CELSR1	A → G (rs4044210)			0.0001	0.0002	0.0010	1.78 (1.25–2.50)		
	AA	938 (94.8)	5137 (97.1)						
	AG	52 (5.3)	154 (2.9)						
	GG	0 (0)	0 (0)						
	Hardy–Weinberg P	0.3961	0.2827						
LLGL2	T → C (rs1671021)			0.0050	0.0011	0.0004	0.74 (0.62–0.87)	0.0887	0.60 (0.32–1.04)
	TT	770 (77.6)	3887 (72.9)						
	TC	208 (21.0)	1329 (24.9)						
	CC	14 (1.4)	119 (2.2)						
	Hardy–Weinberg P	0.9912	0.6672						
RUVBL2	C → T (rs1062708)			0.0277	0.0176	0.6857	0.97 (0.83–1.13)	0.0374	0.82 (0.68–0.99)
	CC	313 (31.6)	1564 (29.6)						
	CT	499 (50.4)	2578 (48.7)						
	TT	178 (18.0)	1149 (21.7)						
	Hardy–Weinberg P	0.3924	0.1564						

OR, odds ratio; CI, confidence interval. Multivariable logistic regression analysis was performed with adjustment for age, sex, BMI, smoking status, and the prevalence of hypertension, diabetes mellitus, and hypercholesterolemia.

[a] Numbers in parentheses are percentages.

Table 5
Relation of *CELSR1* haplotypes to ischemic stroke in all subjects.

Haplotype	Overall frequency	Frequency		Chi-square	Permutation
		Ischemic stroke	Controls	*P* value	*P* value
A–A	0.9836	0.9737	0.9854	1.55×10^{-4}	<0.0001
G–G	0.0154	0.0253	0.0136	1.07×10^{-4}	<0.0001
A–G	9.55×10^{-4}	1.01×10^{-3}	9.45×10^{-4}	0.9314	0.691

Haplotypes consist of the A→ G (rs6007897) and A→ G (rs4044210) polymorphisms of *CELSR1*.

We next examined the relation of three tag SNPs in exons — rs1671021 (T→ C, Phe479Leu) of *LLGL2*, rs6007897 (A→ G, Thr2268Ala) of *CELSR1*, and rs1062708 (C→ T, Leu205Leu) of *RUVBL2* — as well as a nonsynonymous SNP [rs4044210 (A→ G, Ile2107Val) of *CELSR1*] to ischemic stroke in subject panels B and C. The chi-square test revealed that rs6007897 (A→ G, Thr2268Ala) and rs4044210 (A→ G, Ile2107Val) of *CELSR1* and rs1671021 (T→ C, Phe479Leu) of *LLGL2* were significantly (*P*< 0.05) associated with ischemic stroke in terms of both genotype distribution and allele frequency in subject panel B (Table 2). The chi-square test revealed that rs6007897 and rs4044210 of *CELSR1* were also significantly (*P*< 0.05) associated with ischemic stroke in terms of both genotype distribution and allele frequency in subject panel C (Table 3).

The relation of the three tag and one nonsynonymous SNPs to ischemic stroke was also examined in all subjects (combined subject panels A, B, and C). Evaluation of genotype distributions and allele frequencies by the chi-square test revealed that the four polymorphisms were all significantly (*P*< 0.05) associated with ischemic stroke (Table 4). Multivariable logistic regression analysis with adjustment for age, sex, BMI, smoking status, and the prevalence of hypertension, diabetes mellitus, and hypercholesterolemia also revealed that rs6007897 and rs4044210 of *CELSR1* and rs1671021 of *LLGL2* (dominant models) as well as rs1062708 of *RUVBL2* (recessive model) were significantly (*P*< 0.05) associated with ischemic stroke (Table 4). The variant *G* (Ala) allele of rs6007897 and *G* (Val) allele of rs4044210 of *CELSR1* were risk factors for ischemic stroke, whereas the variant *C* (Leu) allele of rs1671021 of *LLGL2* and the *T* allele of rs1062708 of *RUVBL2* were protective against this condition. The genotype distributions of these four SNPs were in Hardy–Weinberg equilibrium among both subjects with ischemic stroke and controls (Table 4).

We performed a stepwise forward selection procedure to examine the effects of genotypes for the four polymorphisms of *CELSR1*, *LLGL2*, and *RUVBL2* as well as of age, sex, BMI, smoking status, and the prevalence of hypertension, diabetes mellitus, and hypercholesterolemia on ischemic stroke (Supplementary Table 10). Diabetes mellitus, hypertension, BMI, smoking, sex, *LLGL2* genotype (dominant model), *CELSR1* genotype (rs6007897, dominant model), age, and *RUVBL2* genotype (recessive model), in descending order of statistical significance, were significant (*P*< 0.05) and independent determinants of ischemic stroke.

Given that rs6007897 and rs4044210 of *CELSR1* were in LD, we performed haplotype analysis for these SNPs. Such analysis revealed that the frequency of the major haplotype, A (rs6007897)–A (rs4044210), was significantly lower, whereas that of the minor haplotype G–G was significantly higher, in subjects with ischemic stroke than in controls (Table 5).

Finally, we examined the relations of the three tag and one nonsynonymous polymorphisms to the prevalence of hypertension, type 2 diabetes mellitus, hypercholesterolemia, or obesity (Supplementary Table 11). Comparison of genotype distributions or allele frequencies of the four polymorphisms between all cases and controls by the chi-square test revealed that rs6007897 and rs4044210 of *CELSR1* as well as rs1062708 of *RUVBL2* were significantly (*P*< 0.05) associated with type 2 diabetes mellitus. Multivariable logistic regression analysis with adjustment for

age, sex, BMI, and smoking status revealed that rs6007897 and rs4044210 of *CELSR1* were significantly associated with type 2 diabetes mellitus with the *G* allele of both polymorphisms representing risk factors for this condition. The rs1062708 of *RUVBL2* was also associated with type 2 diabetes mellitus with the *T* allele protecting against this condition (Supplementary Table 12). The rs6007897 and rs4044210 of *CELSR1* were significantly associated with blood glycosylated hemoglobin content but not with fasting plasma glucose level. The rs1062708 of *RUVBL2* was not related to either parameter (Supplementary Table 13).

4. Discussion

The main cause of ischemic stroke is atherothrombosis, with the principal and treatable risk factors including hypertension, diabetes mellitus, and hypercholesterolemia [11]. In addition to these conventional risk factors, genetic variants are important in the pathogenesis of ischemic stroke [12]. Prediction of the risk for ischemic stroke on the basis of genetic variants would be useful for deciding how aggressively to target the clinical risk factors that are currently amenable to treatment. We have now shown that rs6007897 (A→ G, Thr2268Ala) and rs4044210 (A→ G, Ile2107Val) of *CELSR1* were significantly associated with the prevalence of ischemic stroke in Japanese individuals, with the variant *G* (Ala) and *G* (Val) alleles, respectively, representing risk factors for this condition.

Kubo et al. [13,14] performed a genome-wide association study of ischemic stroke and identified two genes associated with this condition in Japanese individuals. The 1425G→ A (Val374Ile, rs2230500) polymorphism of the protein kinase C, eta gene (*PRKCH*) was significantly associated with lacunar infarction in two independent populations. This polymorphism altered the activity of PRKCH. PRKCH was mainly expressed in vascular endothelial cells and foamy macrophages in human atherosclerotic lesions and its expression was enhanced as the lesion progressed [13]. Four SNPs around the angiotensin II receptor-like 1 gene (*AGTRL1*) were also associated with ischemic stroke. The −154G→ A polymorphism (rs9943582) in the promoter region enhanced the mRNA expression of *AGTRL1* [14]. A population-based cohort study with 14 years of follow-up revealed that functional SNPs of these genes were significantly related to the incidence of ischemic stroke [13,14]. The genes identified by GWAS differed between the previous [13,14] and our studies. Although the reason for the difference remains unclear, cases in the previous studies included subjects with all types of ischemic stroke [13,14], whereas those in our study comprised subjects with atherothrombotic cerebral infarction. Platforms for GWAS also differed between the previous [13,14] and our studies.

Cadherin, epidermal growth factor (EGF) laminin A G-type repeats (LAG) seven-pass G-type receptor 1 (CELSR1) is a member of the flamingo subfamily of cadherin proteins. The flamingo subfamily consists of atypical cadherins that do not interact with catenins. The flamingo cadherins are localized to the plasma membrane and possess nine cadherin domains, seven EGF-like repeats, and two LAG repeats in their ectodomains. They also contain seven transmembrane domains, a characteristic specific to this subfamily. These proteins are thought to function as receptors in contact-

mediated communication, with the cadherin domains acting as homophilic binding regions and the EGF-like domains contributing to cell adhesion and receptor–ligand interactions (Entrez Gene, NCBI) [15–17]. In situ hybridization and reverse transcription-polymerase chain reaction analysis have revealed substantial levels of *Celsr1* expression in the neural tube, brain, lung epithelium, and nascent eyelid of mouse embryos at embryonic day 11.5. In adult mice, *Celsr1* expression was detected in the spinal cord, eye, and brain, predominantly in ependymal cells lining the lateral, third, and fourth ventricles [15]. *Celsr1* is a developmentally regulated, neural-specific gene that plays a role in early embryogenesis (Entrez Gene, NCBI) [15]. The rs4044210 (A → G, Ile2107Val) SNP in exon 17 of *CELSR1* is located in the region of the gene encoding the hormone receptor domain. This extracellular domain contains four conserved cysteine residues that likely form disulfide bridges. Given that this domain is present in a variety of hormone receptors, it may function as a ligand binding domain (Entrez Gene, NCBI).

A SNP (rs646776) at 1p13 near *CELSR2*, which is homologous to *CELSR1*, was shown to be associated with the serum concentration of low density lipoprotein-cholesterol [18] and the prevalence of myocardial infarction [19]. However, rs6007897 and rs4044210 of *CELSR1* were not related to lipid profiles in the present study. Given that identified SNPs of *CELSR1* were related to the prevalence of type 2 diabetes mellitus, the relations of these SNPs to ischemic stroke might be attributable, at least in part, to the effect on diabetes mellitus. Human *CELSR1*, *CELSR2*, and *CELSR3* are planar cell polarity signaling molecules implicated in the regulation of cellular polarity, convergent extension, and invasion. Activation of the planar cell polarity signaling pathway controls tissue polarity and cell movement through the activation of ras homolog gene family, member A (RHOA), mitogen-activated protein kinase 8 (MAPK8), and nemo-like kinase (NLK) [20]. The rs6007897 and rs4044210 of *CELSR1* might influence this signaling pathway, although the effects of these SNPs on protein structure or function remain to be determined. A relation of genetic variants of *CELSR1* to human disease has not previously been demonstrated.

Our results suggest that the lethal giant larvae homolog 2 gene (*LLGL2*) is also a candidate gene for ischemic stroke in Japanese individuals, although the association of rs1671021 (T → C, Phe479Leu) with this condition was not replicated in subject panel C. This gene encodes a protein similar to the lethal (2) giant larvae protein of *Drosophila*, which plays a role in asymmetric cell division, the establishment of epithelial cell polarity, and cell migration [21,22]. The ability of the *Drosophila* protein to localize cell fate determinants is regulated by atypical protein kinase C [22]. The human LLGL2 protein interacts with protein complexes containing atypical protein kinase C and is cortically localized in mitotic cells [22]. Alternative splicing of *LLGL2* transcripts generates multiple mRNA variants that encode different protein isoforms (Entrez Gene, NCBI) [21,22]. The effect of rs1671021 on protein structure or function remains unclear. Variants of *LLGL2* have also not previously been associated with human disease.

There are several limitations in the present study: (i) Given that the study population comprised only Japanese individuals, validation of our findings is required in other ethnic groups. (ii) When the difference to detect was set to 0.01 and alpha to 0.05, the statistical power was not enough in subject panel A (0.2) as compared with that in subject panel B (0.95) or C (0.70). (iii) The number of subjects with ischemic stroke was small in a replication study with community-dwelling individuals (subject panel C). (iv) Given that the minor allele frequencies of *CELSR1* were low (1.5% in rs6007897, 1.6% in rs4044210) in the total population, the clinical implication of the present results was limited. (v) The functional relevance of the identified polymorphisms of *CELSR1* with ischemic stroke remains to be determined.

In conclusion, although the functional relevance of the identified SNPs to protein structure or function was not determined, our present results suggest that *CELSR1* is a susceptibility gene for ischemic stroke in Japanese individuals. Determination of genotypes for the identified SNPs of this gene may prove informative for assessment of the genetic risk for ischemic stroke in Japanese. Validation of our findings will require their replication with independent populations of various ethnic groups.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.atherosclerosis.2009.03.038.

References

[1] Rosamond W, Flegal K, Furie K, et al. Heart disease and stroke statistics—2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 2008;117:e25–146.
[2] Warlow C, Sudlow C, Dennis M, Wardlaw J, Sandercock P. Stroke. Lancet 2003;362:1211–24.
[3] Gretarsdottir S, Thorleifsson G, Reynisdottir ST, et al. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. Nat Genet 2003;35:131–8.
[4] Helgadottir A, Manolescu A, Thorleifsson G, et al. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nat Genet 2004;36:233–9.
[5] Yamada Y, Metoki N, Yoshida H, et al. Genetic risk for ischemic and hemorrhagic stroke. Arterioscler Thromb Vasc Biol 2006;26:1920–5.
[6] Yamada Y, Metoki N, Yoshida H, et al. Genetic factors for ischemic and hemorrhagic stroke in Japanese individuals. Stroke 2008;39:2211–8.
[7] Special report from the National Institute of Neurological Disorders and Stroke. Classification of Cerebrovascular Diseases III. Stroke 1990; 21:637–76.
[8] Itoh Y, Mizuki N, Shimada T, et al. High throughput DNA typing of HLA-A, -B, -C and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. Immunogenetics 2005;57:717–29.
[9] Nakai K, Habano W, Fujita T, et al. Highly multiplexed genotyping of coronary artery disease-associated SNPs using MALDI-TOF mass spectrometry. Hum Mutat 2002;20:133–8.
[10] Kanazawa M, Yoshiike N, Osaka T, et al. Criteria and classification of obesity in Japan and Asia-Oceania. Asia Pac J Clin Nutr 2002;11:S732–7.
[11] Goldstein LB, Adams R, Becker K, et al. Primary prevention of ischemic stroke: a statement for healthcare professionals from the Stroke Council of the American Heart Association. Stroke 2001;32:280–99.
[12] Humphries SE, Morgan L. Genetic risk factors for stroke and carotid atherosclerosis: insights into pathophysiology from candidate gene approaches. Lancet Neurol 2004;3:227–36.
[13] Kubo M, Hata J, Ninomiya T, et al. A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk of cerebral infarction. Nat Genet 2007;39:212–7.
[14] Hata J, Matsuda K, Ninomiya T, et al. Functional SNP in an Sp1-binding site of AGTRL1 gene is associated with susceptibility to brain infarction. Hum Mol Genet 2007;16:630–9.
[15] Hadjantonakis AK, Sheward WJ, Harmar AJ, et al. Celsr1, a neural-specific gene encoding an unusual seven-pass transmembrane receptor, maps to mouse chromosome 15 and human chromosome 22qter. Genomics 1997;45:97–104.
[16] Wu Q, Maniatis T. A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell 1999;97:779–90.
[17] Nollet F, Kools P, van Roy F. Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. J Mol Biol 2000;299:551–72.
[18] Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet 2008;40:189–97.
[19] Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009;41:334–41.
[20] Katoh Y, Katoh M. Comparative integromics on FAT1, FAT2, FAT3 and FAR4. Int J Mol Med 2006;18:523–8.
[21] Müsch A, Cohen D, Yeaman C, et al. Mammalian homolog of Drosophila tumor suppressor lethal (2) giant larvae interacts with basolateral exocytic machinery in Madin-Darby canine kidney cells. Mol Biol Cell 2002;13:158–68.
[22] Yasumi M, Sakisaka T, Hoshino T, et al. Direct binding of Lgl2 to LGN during mitosis and its requirement for normal cell division. J Biol Chem 2005;280:6761–5.

PLoS one

RNA Editing Genes Associated with Extreme Old Age in Humans and with Lifespan in *C. elegans*

Paola Sebastiani[1]*, Monty Montano[2]*, Annibale Puca[3], Nadia Solovieff[1], Toshio Kojima[4], Meng C. Wang[5], Efthymia Melista[6], Micah Meltzer[2], Sylvia E. J. Fischer[5], Stacy Andersen[7], Stephen H. Hartley[1], Amanda Sedgewick[8], Yasumichi Arai[9], Aviv Bergman[10], Nir Barzilai[11], Dellara F. Terry[7], Alberto Riva[12], Chiara Viviani Anselmi[3], Alberto Malovini[3], Aya Kitamoto[4], Motoji Sawabe[13], Tomio Arai[13], Yasuyuki Gondo[14], Martin H. Steinberg[6], Nobuyoshi Hirose[9], Gil Atzmon[11], Gary Ruvkun[5], Clinton T. Baldwin[6], Thomas T. Perls[7]

1 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, 2 Department of Medicine Sections of Infectious Diseases, Boston University School of Medicine, Boston, Massachusetts, United States of America, 3 Department of Genetics, IRCCS Multimedica, Milan, Italy, 4 Computational Systems Biology Research Group, Advanced Science Institute, RIKEN, Yokohama, Kanagawa, Japan, 5 Department of Genetics, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, 6 Center for Human Genetics, Boston University School of Medicine, Boston, Massachusetts, United States of America, 7 Geriatrics Section, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, 8 Department of Bioengineering, Boston University School of Engineering, Boston, Massachusetts, United States of America, 9 Division of Geriatric Medicine, Department of Internal Medicine, Keio University School of Medicine, Tokyo, Japan, 10 Department of Systems & Computational Biology, Albert Einstein College of Medicine, Bronx, New York, United States of America, 11 Institute of Aging Research and Diabetes Research Center, Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, New York, United States of America, 12 Department of Molecular Genetics and Microbiology, University of Florida at Gainesville, Florida, United States of America, 13 Department of Pathology, Tokyo Metropolitan Geriatric Hospital, Tokyo, Japan, 14 Clinical Thanatology and Geriatric Behavioral Science, Graduate School of Human Sciences, Osaka University, Suita, Osaka, Japan

Abstract

Background: The strong familiality of living to extreme ages suggests that human longevity is genetically regulated. The majority of genes found thus far to be associated with longevity primarily function in lipoprotein metabolism and insulin/IGF-1 signaling. There are likely many more genetic modifiers of human longevity that remain to be discovered.

Methodology/Principal Findings: Here, we first show that 18 single nucleotide polymorphisms (SNPs) in the RNA editing genes *ADARB1* and *ADARB2* are associated with extreme old age in a U.S. based study of centenarians, the New England Centenarian Study. We describe replications of these findings in three independently conducted centenarian studies with different genetic backgrounds (Italian, Ashkenazi Jewish and Japanese) that collectively support an association of *ADARB1* and *ADARB2* with longevity. Some SNPs in *ADARB2* replicate consistently in the four populations and suggest a strong effect that is independent of the different genetic backgrounds and environments. To evaluate the functional association of these genes with lifespan, we demonstrate that inactivation of their orthologues *adr-1* and *adr-2* in *C. elegans* reduces median survival by 50%. We further demonstrate that inactivation of the argonaute gene, *rde-1*, a critical regulator of RNA interference, completely restores lifespan to normal levels in the context of *adr-1* and *adr-2* loss of function.

Conclusions/Significance: Our results suggest that RNA editors may be an important regulator of aging in humans and that, when evaluated in *C. elegans*, this pathway may interact with the RNA interference machinery to regulate lifespan.

* E-mail: sebas@bu.edu (PS); mmontano@bu.edu (MM)

⑨ These authors contributed equally to this work.

¶ These authors also contributed equally to this work.

Introduction

Exceptional longevity (EL) in humans, defined as living to extreme old ages such as 100 years and older, is strongly familial [1–8] and the factors that facilitate such exceptional survival have broad public health significance including a marked delay in age-related disability [9–11] and certain age-related diseases [12–14]. Genetically, exceptional longevity is presumed to be a complex trait [15–19]. Several candidate gene association studies have been successful in discovering longevity-associated

genes in humans. However, these variants have been mainly related to lipoprotein metabolism [20–22], FOXO proteins [23,24], and insulin/IGF-1 signaling [25] It is likely that many more genetic modifiers of human aging have yet to be discovered [25].

In this study, we investigate two genes in the A(adenosine) to I (inosine) RNA editing pathway, a post-transcriptional process by which adenosine residues are converted to inosine resulting in a change in gene expression or protein function. Targets of RNA editing include a large number of genes as well as micro RNA. Thus, it is not surprising that such a non-specific cellular process would be involved in a general maintenance of cellular health and lifespan. However, such an implication has not been previously demonstrated.

Here, we first report significant association of these genes with EL in four centenarian studies that include the New England Centenarian Study (NECS), with more than 1,500 US individuals of primarily North European ancestry, aged between 90 and 119 years; the Southern Italian Centenarian Study (SICS) –a study of nonagenarians and centenarians from a closed population of Cilento, South Italy; the Ashkenazi Jewish Centenarian Study (AJCS), a study of approximately 300 nonagenarians and centenarians from a founder population of North Eastern European background, all resident in the US; and the Japanese Centenarian Study (JCS), a study of Japanese centenarians that has focused on "semi-supercentenarians" subjects living past 105 years [26]. The characteristics of the four populations allow us to assess the robustness of the associations to varying genetic background and environment.

We further evaluate the functional significance of the RNA editing candidate genes in *C. elegans* lifespan studies and show that silencing orthologs of these genes reduces median survival by 50%. We also show that life span is fully restored by additional knockdown of an RNA interference gene, supporting the functional role of these genes in determining lifespan and implicating a novel axis for future aging studies.

Results and Discussion

Selection of Candidate Genes

We selected the two genes to study for multiple reasons. First, in a preliminary genome wide screening using pooled DNA samples from approximately 130 male centenarians and 130 younger male controls from the NECS [27], we identified several single nucleotide polymorphisms (SNPs) in the RNA-editing genes *ADARB1* (21q22.3), and *ADARB2* (10p15.3) that were associated with extreme old age. *ADARB1* exhibited the strongest evidence for genetic association with 5 SNPs that met genome-wide significance, with the posterior odds of allelic association >1,500 [27]. The probability of these 5 SNPs simultaneously associated under the null hypothesis of no association was 10^{-13} based upon a hyper-geometric distribution. Second, this gene lies in chromosome 21q21 and trisomy 21 (Down syndrome) resembles accelerated aging, with premature age-related changes including in the skin and hair, increased frequency of premature cataracts, hearing loss, menopause and Alzheimer's disease [28] suggesting that genes in chromosome 21 could affect lifespan. Third, among the top genes identified from the preliminary genetic screen, RNA editing represents a general cellular process that might be expected to improve cellular health; and RNA editing activity has been associated with innate immune response [29,30] and age-related syndromes that include dementia and amyotrophic lateral sclerosis (ALS) [31].

Subjects Selected for the Association Study

From the NECS, genotype data were obtained from 281 males, aged 96–114 years and 596 females, aged 100–119 years. We selected cutoff ages of 96 years for males and 100 years for females of the NECS to focus on the extreme top 1% survival based on the U.S. Social Security Administration cohort life table (http://www.ssa.gov/OACT/NOTES/as116/as116LOT.html). NECS referent cohort subjects consisted of 270 spouses of centenarian offspring and children of parents who died at the mean age of 73 years (average life expectancy for the parents' birth cohort). Additional referent subjects were selected from the Illumina iControlDB database using genome-wide genetic matching as detailed in the methods (n = 1635). Note that approximately 100 male centenarians included in the pooling-based genome screening overlap with this second set. Given that the overlap is relatively small (<10%) and that the subsequent analysis uses a different analytic approach (genotype data from individual subjects), we do not think the overlap is a significant concern.

From the SICS, we used genotype data from 271 males, ages 90–108 years and 188 females, aged 90–109 years (total = 459). Data from 200 male and 132 female SICS referent cohort subjects aged 18–48 years were used in this analysis. From the AJCS, genotype data were obtained from 299 oldest subjects (108 males aged 95 and older and 191 females aged 99 and older) and 269 younger referent cohort subjects (spouses of the offspring of centenarians, aged 85 and younger, without evidence of parental longevity). Four hundred and seventy oldest old subjects (82 males aged 100–110 years and 388 females, aged 100–116 years) and 538 referent cohort subjects (randomly selected Japanese subjects, aged 19–89 years) constituted the Japanese association study. Table 1 reports further summaries of subjects' characteristics. Ages of the extreme old were validated with birth certificates (in the case

Table 1. Study Subjects characteristics.

	Males	Females	All	Males	Females	All
	NECS oldest old			**NECS controls**		
Sample Size	281	596	877	149	121	270
Median Age	102	103	103	75	74	75
Age Range	96–114	100–119	96–119	58–85	53–85	53–85
	SICS oldest old			**SICS controls**		
Sample Size	271	188	459	200	132	332
Median Age	94	98	96	34	32	33
Age Range	90–109	90–109	90–109	18–48	18–48	18–48
	NA			**Illumina controls**		
Sample Size				418	1217	1635
Median Age				47	46	47
Age Range				30–75	30–75	30–75
	AJCS oldest old			**AJCS controls**		
Sample Size	108	191	299	118	151	269
Median Age	99	101	100	77	73	73
Age Range	95–108	99–112	95–112	54–85	46–85	46–85
	JCS oldest old			**JCS controls**		
Sample Size	82	388	470	178	360	538
Median Age	104	106	106	21	72	69
Age Range	100–111	100–116	100–116	19–89	19–89	19–89

Reported are summaries of the last contact ages.
doi:10.1371/journal.pone.0008210.t001

of the JCS, the Basic Resident Registration Card). All subjects were enrolled by studies with Institutional Review Board approval and oversight.

Association of ADARB1 with Exceptional Longevity

We examined the associations of 31 SNPs in *ADARB1* in the NECS and SICS samples using recessive and dominant models with Bayesian logistic regression [32]. The details of the statistical analysis are in the Methods and the significant results are summarized in Table 2 that provides the physical positions and allele frequencies derived from the HapMap for these SNPs, and Table 3 (rows 14–15). Five SNPs in *ADARB1* are strongly associated with extreme old age in the NECS, and the association of SNP rs414743 remains significant even after imposing stringent corrections for multiple comparisons (Bayesian significance $< 0.05/145 = 0.00035$ where 145 is the overall number of SNPs included in this analysis). The five SNPs tag one region of strong linkage disequilibrium (LD) of the gene (Figure 1). None of the SNPs reached statistical significance in the SICS although the three SNPs rs2838809, rs2838810 and rs2838816 exhibited consistent associations in terms of odds ratios and allele frequencies and, when the NECS and SICS data were combined, the three SNPs remained statistically significant. These three SNPs have extreme minor allele frequencies in the NECS centenarians (MAF<0.01), while the allele frequencies in the controls are very close to referent allele frequencies from HapMap (Table 3). Figure 2 displays the scatter plot of genotype intensities generated from BeadStudio that rules out genotyping

errors thus suggesting that these are real associations and not artifacts.

To further test the generalizability of these results to other independent groups, we evaluated these associations in the AJCS and the JCS, using a combination of proxy SNPs typed with the Affymetrix platforms and SNPs in Table 3 typed with more traditional techniques (See methods). Table 4 summarizes the results of the replication study of 4 of the 5 SNPs and one additional proxy SNP is in Table 5. None of these SNPs in *ADARB1* replicates the results in the NECS and SICS samples although the significant association of the SNP rs17004734 that is within 2Kb from rs414743 is consistent with the presence of longevity associated variants in the region. Because the SNPs used in the NECS and SICS are chosen to best capture the genetic variations of Caucasians from the HapMap, they may not be the correct choice for this founder population and indeed Figure 1 shows a different pattern of LD in *ADARB1* in the AJCS subjects. All *ADARB1* SNPs in Table 3 and two additional proxy SNPs were genotyped in the JCS subjects and Tables 6 and 7 summarize the results. The last three SNPs in Table 6 show effects that are consistent with the NECS and SICS subjects but do not reach statistical significance, even when the data from the three studies are aggregated. Figure 3 shows the posterior densities of the ORs for the three rare SNPs that are suggestive of association but would need much larger sample sizes to reach statistical significance. The association of two proxy SNPs for rs2838816 in Table 7 is again consistent with the presence of longevity

Table 2. Summary of selected SNPs.

Row	SNP	Chr	position	Risk versus referent alleles	CEPH	JPT
1	rs10903420	10	1333726	AA v AG/GG	0.327	0.058
2	rs1007147	10	1341088	AA v AG/GG	0.312	0.091
3	rs2805562	10	1357514	AA v AG/GG	0.15	0.058
4	rs884949	10	1361610	AA v AC/CC	0.124	0
5	rs2805533	10	1374633	AA/AG v GG	0.77	0.92
6	rs2387653	10	1397826	AA v AG/GG	0.097	0
7	rs2805535	10	1450432	AA v AG/GG	0.699	0.151
8	rs2805543	10	1454892	AA v AG/GG	0.46	0.105
9	rs3898610	10	1474759	AA v AG/GG	0.841	0.686
10	rs1533484	10	1481339	AA/AG v GG	0.442	0.791
11	rs2676192	10	1495474	AA v AG/GG	0.301	0.419
12	rs2387952	10	1657365	AA v AG/GG	0.69	0.616
13	rs17294019	10	1659347	AA/AG v GG	0.196	0.012
14	rs3788157	21	45335136	AA v AG/GG	0.63	0.65
15	rs414743	21	45336503	AA/AG v GG	0.47	0.31
16	rs2838809	21	45445866	AA v AG/GG	0.009	0
17	rs2838810	21	45447751	AA/AG v GG	1	1
18	rs2838816	21	45454470	AA v AG/GG	0.009	0

List of 18 SNPs —13 in the gene *ADARB2* (10p15.3) and 5 in the gene *ADARB1* (21q22.3) — that are associated with exceptional longevity with either dominant or recessive models. The table reports a sequential number for easy identification in the other tables and figures (column 1), the SNP identifier from dbSNP (column 2), chromosome and physical position from the human genome NCBI Build 36.3 (columns 3 and 4), the risk versus referent alleles that were associated with exceptional longevity using dominant and recessive models (column 5), the frequencies of the risk allele in the HapMap CEPH and JPT. Note that several of alleles in the Japanese group have allele frequencies that change substantially from the CEPH, and the SNPs in rows 4 and 6 become monomorphic.
doi:10.1371/journal.pone.0008210.t002

Table 3. SNPs in *ADARB2* (10p15.3) and *ADARB1* (21q22.3) that are associated with exceptional longevity in NECS and SICS subjects.

Row	SNP	Risk versus referent alleles	CEPH	NECS (877 oldest old, 1808 controls)			SICS (459 oldest old, 429 controls)			NECS+SICS		
				OR	Bayes sig	p(A)	OR	Bayes sig	p(A)	OR	Bayes sig	p(A)
ADARB2												
1	rs10903420	AA v AG/GG	0.327	1.28(1.07;1.53)	0.0048	0.28/0.23	1.14(0.84;1.54)	0.1212	0.28/0.25	1.25(1.07;1.45)	0.0020	0.28/0.24
2	rs1007147	AA v AG/GG	0.312	1.35(1.13;1.64)	0.0015	0.27/0.22	1.25(0.92;1.66)	0.2040	0.29/0.25	1.34(1.15;1.59)	0.0003	0.27/0.22
3	rs2805562	AA v AG/GG	0.15	1.22(0.96;1.54)	0.0500	0.15/0.12	1.38(0.97;1.97)	0.0384	0.20/0.15	1.32(1.09;1.59)	0.0022	0.16/0.13
4	rs884949	AA v AC/CC	0.124	1.19(0.92;1.53)	0.0911	0.12/0.10	1.31(0.87;1.96)	0.0978	0.14/0.11	1.24(1.01;1.54)	0.0211	0.12/0.10
5	rs2805533	AA/AG v GG	0.77	0.91(0.75;1.12)	0.1904	0.78/0.81	0.83(0.60;1.14)	0.1217	0.76/0.79	0.86(0.72;1.02)	0.0381	0.79/0.81
6	rs2387653	AA v AG/GG	0.097	1.17(0.91;1.50)	0.1053	0.14/0.12	1.12(0.78;1.62)	0.2699	0.17/0.16	1.21(0.99;1.49)	0.0343	0.15/0.13
7	rs2805535	AA v AG/GG	0.699	1.36(1.02;1.83)	0.0249	0.75/0.69	1.31(0.97;1.79)	0.0395	0.72/0.66	1.37(1.11;1.68)	0.0017	0.74/0.68
8	rs2805543	AA v AG/GG	0.46	1.23(1.04;1.45)	0.0055	0.54/0.49	1.36(1.03;1.76)	0.0156	0.51/0.43	1.24(1.08;1.42)	0.0008	0.53/0.48
9	rs3898610	AA v AG/GG	0.841	1.42(1.11;1.77)	0.0015	0.88/0.83	1.19(0.80;1.74)	0.1973	0.87/0.85	1.36(1.12;1.67)	0.0015	0.87/0.84
10	rs1533484	AA/AG v GG	0.442	0.81(0.70;0.97)	0.0109	0.35/0.40	0.79(0.63;1.01)	0.0428	0.43/0.49	0.86(0.75;0.98)	0.0147	0.38/0.42
11	rs2676192	AA v AG/GG	0.301	0.81(0.67;0.98)	0.0096	0.26/0.30	0.95(0.74;1.31)	0.3869	0.26/0.27	0.83(0.71;0.96)	0.0086	0.26/0.30
12	rs2387952	AA v AG/GG	0.69	1.37(1.15;1.66)	0.0004	0.75/0.69	1.05(0.80;1.37)	0.3484	0.66/0.65	1.20(1.04;1.40)	0.0073	0.72/0.68
13	rs17294019	AA/AG v GG	0.196	0.69(0.55;0.86)	0.0005	0.14/0.19	0.80(0.59;1.10)	0.0826	0.22/0.27	0.78(0.65;0.93)	0.0030	0.17/0.21
ADARB1												
14	rs3788157	AA v AG/GG	0.63	1.23(1.05;1.45)	0.0075	0.65/0.62	0.90(0.68;1.19)	0.2318	0.65/0.67	1.16(1.00;1.34)	0.0204	0.65/0.62
15	rs414743	AA/AG v GG	0.47	0.72(0.61;0.85)	$<10^{-5}$	0.40/0.48	1.08(0.83;1.45)	0.2689	0.49/0.46	0.83(0.72;0.96)	0.0053	0.43/0.48
16	rs2838809	AA v AG/GG	0.009	0.27(0.05;0.93)	0.0096	0.003/0.009	0.58(0.15;1.97)	0.1973	0.011/0.016	0.46(0.19;1.08)	0.0297	0.005/0.01
17	rs2838810	AA/AG v GG	1	3.71(1.12;19.29)	0.0136	0.997/0.990	2.29(0.63;10.33)	0.1088	0.991/0.984	2.73(1.11;8.25)	0.0136	0.995/0.99
18	rs2838816	AA v AG/GG	0.009	0.28(0.05;0.93)	0.0116	0.003/0.009	0.58(0.15;1.98)	0.1934	0.011/0.016	0.48(0.18;1.04)	0.0309	0.005/0.01

The first 13 SNPs are in the gene *ADARB2* (10p15.3) and the last five SNPs in the gene *ADARB1* (21q22.3). Columns 1–4 provide details of the SNPs as in table 2. Columns 5–7 report the results of the association in the NECS and referent subjects based on Bayesian logistic regression of dominant and recessive models. Specifically, column 5 reports the Bayesian estimate of the odds ratio and 95% credible interval within brackets, columns 6 reports the Bayes significance that is defined as $1-p(OR>1)$ when the posterior estimate of the OR is >1 and $1-p(OR<1)$ when the posterior estimate of the OR is <1. This number is the posterior probability of the null hypothesis $OR\leq1$ when we estimate $OR>1$ (or $OR\geq1$ when we estimate $OR<1$) so small values provide strong evidence against the null hypothesis and it is similar to the Bayes p-value proposed by Althman. Column 7 reports the posterior probability of the risk allele in cases and controls. Highlighted in bold are the significant associations (Bayes significance <0.05). Columns 8–10 report the replications in the SICS and highlighted in bold are the 4 SNPs that are significant in this analysis. Columns 11–12 report the results of the analysis when the data from the two studies are aggregated. Although only 15 SNPs reach statistical significance in the NECS and only 4 SNPs in the SICS, all 18 have consistent effects and when data of the two studies are aggregated, all SNPs remain significant and the significance of 11 of them becomes stronger.
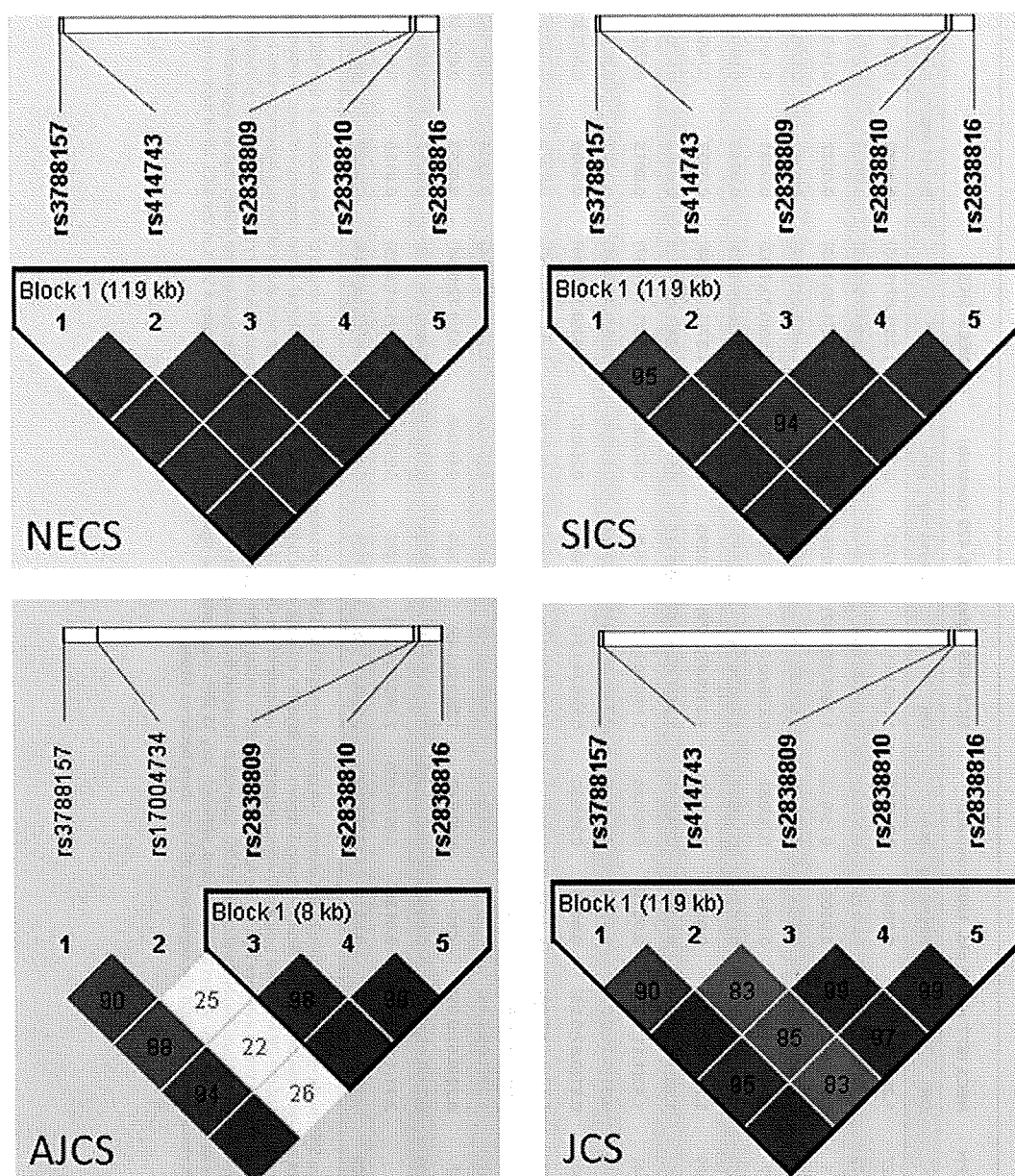doi:10.1371/journal.pone.0008210.t003

Figure 1. Pattern of LD among the SNP in *ADARB1* (chromosome 21) that are associated with exceptional longevity. The four plots display the pattern of LD captured by the SNPs associated with exceptional longevity in *ADARB1* (chromosome 21) using data from the NECS, SICS, AJCS and JCS. The intensity of red represents the strength of LD measured by D'.
doi:10.1371/journal.pone.0008210.g001

associated variants in this gene that may not be captured by our SNP selection in the Japanese population.

Association of ADARB2 with Exceptional Longevity

We examined the associations of 114 SNPs in *ADARB2* in the NECS and SICS samples using the same recessive and dominant models. Ten SNPs were strongly associated with extreme old age in the NECS, and one remains significant even after correcting for multiple comparisons (SNPs rs2387952, Bayesian significance 0.0004~0.05/145). Four of these significant SNPs (rs2805562; rs2805533; rs2805543; and rs1533484) were also replicated in the SICS (Bayes significance <0.05) (Table 3). The remaining six SNPs did not reach statistical significance in the SICS but did

exhibit consistent associations in terms of odds ratios and allele frequencies and combining data from the NECS and SICS made these ten associations even stronger plus an additional three other SNPs became statistically significant. These SNPs tag a region of approximately 160Kb in *ADARB2* that includes two blocks of LD (Figure 4).

Tables 4 and 5 summarize the results of the replication of 10 of these SNPs in AJCS subjects. Two of the SNPs reach statistical significance in this set (rows 3 and 13) but one has opposite effects compared to the NECS and SICS (rows 13). However, SNPs in rows 1, 5–9, 11 and 12 have effects that are consistent with the NECS and SICS and, when the data of the 3 studies are aggregated, they become significant (columns 12–14). The SNPs in
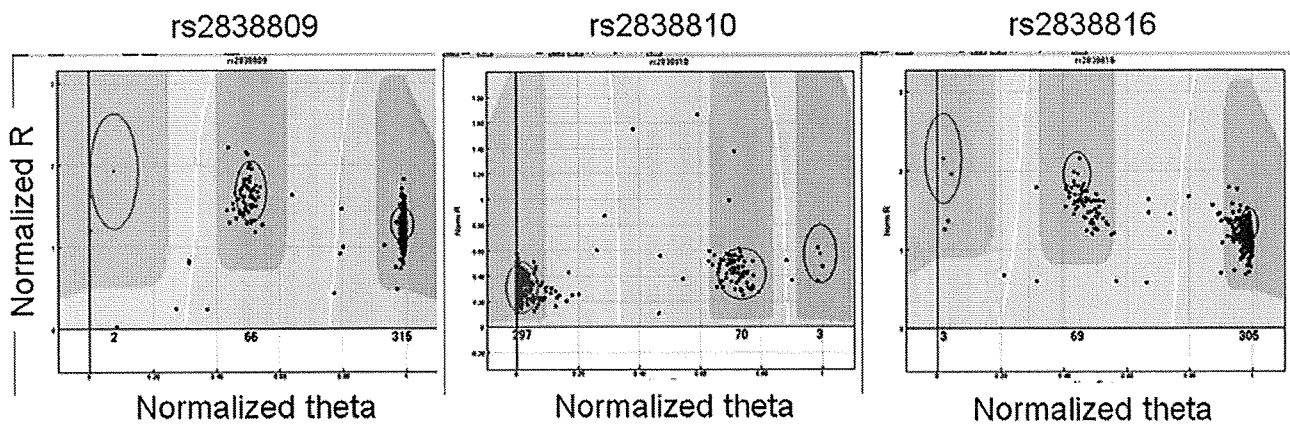
Figure 2. Result of genotype cluster algorithm from BeadStudio. The three plots show the normalized intensities in polar coordinate and the cluster definition from BeadStudio for NECS subjects. The clear separation suggests that the genotype calls are robust.
doi:10.1371/journal.pone.0008210.g002

rows 2, 4, and 11 were not typed but proxy SNPs typed with the Affymetrix 6.0 array confirm strong associations of variants in the same region with EL. These SNPs are summarized in Table 5 and are a good proxy for the SNPs originally typed in the NECS and SICS as shown by their proximity in terms of physical distance <10kb and linkage disequilibrium measured by D' and r^2.

Only SNP rs2805533 reached statistical significance in the JCS set, with an effect that was consistent with the observed effect in the NECS, SICS and AJCS samples (Table 6). The SNP rs1533484 was borderline significant (Bayesian significance ~0.06) and demonstrates consistent effects with the NECS and SICS results, but the allele frequencies are substantially different. The SNP rs10903420 was also significant but with an opposite effect compared to the NECS, SICS and AJCS subject sets. Note however the substantial differences in both allele frequencies and pattern of LD that may explain the different patterns of associations in this ethnically very distinct sample. Genotype data of additional SNPs in Table 7 provide further evidence for the existence of variants in the region between physical positions 1340K and 1500K of chromosome 10 that are associated with exceptional longevity.

Age Trends

For some SNPs, a clear monotonic pattern associated with increasing age was observed (Figure 5). This monotonic pattern is consistent with a genetic effect that results from alleles positively associated with EL becoming more frequent in older individuals, while alleles that are negatively associated with EL become less frequent. This pattern has also been observed for ApoE alleles [33] and is consistent with the phenomenon of demographic selection [34]. These increases in allele frequencies with age also illustrate the increasing gain of power conferred by studying centenarians and even more so, subjects age 105+ years, in genetic studies of exceptional survival.

Functional Studies

To go beyond statistical association, we chose to investigate the possible functional role of these genes in regulating lifespan by evaluating their influence on lifespan in the nematode, *C. elegans*, a robust model organism for candidate lifespan gene discovery. The A-to-I RNA editing gene family and their enzymatic editing activity has been well conserved in a broad array of species including humans, mice, flies, zebrafish, xenopus and notably, *C.*

elegans [31,35]. For lifespan analysis, we focused on *C. elegans*, which has two orthologues of *ADARB1* and *ADARB2* with RNA editing activity, *adr-1* and *adr-2* [36] (see phylogenetic tree in Figure S1). Because *C. elegans* has an average lifespan of approximately 20 days, the influence of candidate genes on lifespan can be readily tested. Therefore, to evaluate whether *adr* loss-of-function influences *C. elegans* lifespan, we monitored the lifespan of single and double mutants of *adr-1(gv6)* and *adr-2(gv42)*. Both of these alleles are deletions that remove at least a third of the coding sequence and are presumed null alleles [36] (Figure 6a and Figure S2). Strains carrying mutations in *adr-1* and *adr-2* displayed a shorter lifespan than the wild-type control N2 worms (log-rank test $p < 10^{-8}$). Remarkably, aside from the decline in lifespan, there were no other obvious defects, in contrast with gain-of-function studies that noted lethality in Drosophila [37].

The insulin-like growth factor (IGF) pathway is a well known lifespan regulatory axis in worms [38], flies [39,40], mice [41], and humans [42]. Inactivation of the *C. elegans* insulin like receptor gene *daf-2* by dsRNA increased lifespan of both the wild type N2 worms, confirming previous studies. Notably, knockdown of *daf-2* by dsRNA also increased lifespan in the background of *adr-1* and *adr-2* null mutations, resulting in a lifespan phenocopy similar to the wild type N2 worms (Figure 6a). Similar extensions of lifespan were seen with single *adr-1* or *adr-2* mutants, (Figure S2). These data suggest that IGF axis mediated increases in lifespan due to *daf-2* remain active in the presence of *adr-1* and *adr-2* background genotypes, but with less potency than in a N2 wild type background. From these data, we cannot exclude the possibility that knockdown of RNA editing genes in *adr-1* or *adr-2* mutants results in increased RNAi activity. In this scenario, increased RNAi might target genes downstream of *daf-16* (seven *daf-16* gene targets have been identified in comparative analysis of daf-2 versus *daf-2::daf-16* strains [43]), thereby reducing the potency of *daf-2* dependent increases in lifespan.

In a previous study, mutations in both *adr-1* and *adr-2* resulted in increased GFP reporter transgene silencing, suggesting that declines in ADAR function are associated with an increase in RNA interference (RNAi), which would account for the GFP silencing in those experiments. When the argonaute gene *rde-1*, which is essential for RNA induced silencing complex (RISC) formation, was introduced into *adr-1;adr-2* worms containing the transgenes, the increased GFP silencing due to ADAR knockdown was suppressed [44]. This suggests cross-regulation between RNA editors and RNA interference that is further supported by results

Table 4. Replication in the AJCS subject set.

	Row	SNP	Risk versus referent alleles	CEPH	NECS+SICS			AJCS (299 oldest old, 269 controls)			NECS+SICS+AJCS		
					OR	Bayes sig	p(A)	OR	Bayes sig	p(A)	OR	Bayes sig	p(A)
ADARB2	1	rs10903420	AA v AG/GG	0.327	1.25(1.07;1.45)	0.0020	0.28/0.24	1.37(0.87;2.18)	0.0861	0.27/0.21	1.26(1.09;1.45)	0.0010	0.28/0.24
	2	rs1007147	AA v AG/GG	0.312	1.34(1.15;1.59)	0.0003	0.27/0.22						
	3	rs2805562	AA v AG/GG	0.15	1.32(1.09;1.59)	0.0022	0.16/0.13	1.88(1.10;3.22)	0.0095	0.21/0.17	1.39(1.16;1.65)	<10−5	0.17/0.13
	4	rs884949	AA v AC/CC	0.124	1.24(1.01;1.54)	0.0211	0.12/0.10						
	5	rs2805533	AA/AG v GG	0.77	0.86(0.72;1.02)	0.0381	0.79/0.81	0.73(0.45;1.18)	0.0955	0.76/0.81	0.84(0.72;0.98)	0.0129	0.78/0.81
	6	rs2387653	AA v AG/GG	0.097	1.21(0.99;1.49)	0.0343	0.15/0.13	1.40(0.83;2.38)	0.1030	0.20/0.15	1.27(1.05;1.53)	0.0082	0.17/0.13
	7	rs2805535	AA v AG/GG	0.699	1.37(1.11;1.68)	0.0017	0.74/0.68	1.27(0.82;1.95)	0.1476	0.81/0.77	1.28(1.06;1.54)	0.0050	0.73/0.70
	8	rs2805543	AA v AG/GG	0.46	1.24(1.08;1.42)	0.0008	0.53/0.48	1.26(0.85;1.87)	0.1217	0.51/0.45	1.22(1.08;1.39)	0.0001	0.52/0.47
	9	rs3898610	AA v AG/GG	0.841	1.36(1.12;1.67)	0.0015	0.87/0.84	0.63(0.35;1.16)	0.0685	0.87/0.91	1.29(1.07;1.55)	0.0035	0.87/0.84
	10	rs1533484	AA/AG v GG	0.442	0.86(0.75;0.98)	0.0147	0.38/0.42						
	11	rs2676192	AA v AG/GG	0.301	0.83(0.71;0.96)	0.0086	0.26/0.30	0.77(0.52;1.15)	0.0986	0.27/0.32	0.83(0.72;0.95)	0.0037	0.26/0.30
	12	rs2387952	AA v AG/GG	0.69	1.20(1.04;1.40)	0.0073	0.72/0.68	0.95(0.61;1.44)	0.3977	0.67/0.69	1.17(1.01;1.34)	0.0146	0.71/0.68
	13	rs17294019	AA/AG v GG	0.196	0.78(0.65;0.93)	0.0030	0.17/0.21	1.42(0.95;2.13)	0.0451	0.31/0.24			
ADARB1	14	rs3788157	AA v AG/GG	0.63	1.16(1.00;1.34)	0.0204	0.65/0.62	0.73(0.50;1.05)	0.0463	0.58/0.65			
	15	rs414743	AA/AG v GG	0.47	0.83(0.72;0.96)	0.0053	0.43/0.48						
	16	rs2838809	AA v AG/GG	0.009	0.46(0.19;1.08)	0.0297	0.005/0.01	1.19(0.44;3.29)	0.3662	0.040/0.034			
	17	rs2838810	AA/AG v GG	1	2.73(1.11;8.25)	0.0136	0.995/0.99	0.99(0.39;2.53)	0.4951	0.956/0.956			
	18	rs2838816	AA v AG/GG	0.009	0.48(0.18;1.04)	0.0309	0.005/0.01	1.06(0.38;3.06)	0.4573	0.035/0.034			

List of the SNPs in *ADARB2* and *ADARB1* that reach statistical significance in the NECS and SICS and were attempted to be replicated in the AJCS set. The first 7 columns report the details of the SNPs as described in the legend of Table 3. Columns 8–11 report the results for 14 of the 18 SNPs in Table 3 that were genotyped in 255 oldest old and 227 younger controls of the AJCS. Only 3 SNPs reach statistical significance in this set (rows 3, 13 and 14) but two of them have opposite effects compared to the NECS and SICS (rows 13 and 14). However, SNPs in rows 1, 5–9, 11 and 12 have effects that are consistent with the NECS and SICS and, when the data of the 3 studies are aggregated, they become significant (columns 12–14). The SNPs in rows 2, 4, 10 and 11 were not typed because proxy SNPs from an ongoing genome wide association study conducted with the Affymetrix 6.0 array suggest strong associations in the same region.
doi:10.1371/journal.pone.0008210.t004

Table 5. Replication in the AJCS subject set using proxy SNPs.

Row	SNP	Risk versus referent alleles	CEPH	NECS+SICS			Proxy SNPs				AJCS (255 oldest old, 227 controls)			Risk versus referent allele
				OR	Bayes sig	p(A)	SNP	Position	Distance	D'/r²	OR	Bayes sig	p(A)	
2	rs1007147	AA v AG/GG	0.312	1.34(1.15;1.59)	0.0003	0.27/0.22	rs2804097	1352129	11041	0.63/0.3	0.56(0.33;0.91)	0.01	0.76/0.85	AA/AT v TT
4	rs884949	AA v AC/CC	0.124	1.24(1.01;1.54)	0.0211	0.12/0.10	rs10903426	1361386	−224	1.00/0.58	1.62(1.00;2.73)	0.0269	0.23/0.16	CC v CT/TT
15	rs414743	AA/AG v GG	0.47	0.83(0.72;0.96)	0.0053	0.43/0.48	rs17004734	45345886	2383	1.00/0.07	0.29(0.06;1.00)	0.0259	0.95/0.98	AA/AG v GG

Additional proxy SNPs that were genotyped in the AJCS set and tag the same region as shown by their proximity in terms of physical distance <10kb and linkage disequilibrium measured by D' and r² (See also Figure 1).
doi:10.1371/journal.pone.0008210.t005

from Tonkin et al [45], wherein they demonstrated that a mild chemotaxis defect of the adr-1;adr-2 double mutant could be rescued by an rde-1 mutant [45]. Therefore, to evaluate the potential for cross-regulation between RNA editing and RNA interference in the context of lifespan, we evaluated adr-1; adr-2 mediated declines in lifespan in the presence of the RNAi defective strain, rde-1(ne-219). Remarkably, and consistent with previous results, the loss of rde-1 completely restored lifespan declines associated with adr-1; adr-2 loss-of-function (Figure 6b). We interpret these data as expanding the interaction between these two RNA regulatory pathways to include lifespan determination.

Our experiments with C. elegans raise the question as to the precise role for A-to-I RNA editing gene activity in human aging, which remains unknown. However, the demonstration that this family of genes is implicated in the regulation of aging in other organisms warrants validation in other species, particularly humans, and provides a novel regulatory axis for future studies on regulatory pathways that influence the aging process. We speculate that as of yet unidentified ADAR variants delay age associated declines in ADAR activity. We note that a reduction in ADAR enzymatic activity is associated with Dementia, ALS and Alzheimers disease in normal aging individuals [46]. Consistent with this interpretation, analysis of published transcriptional profiles of aging in C. elegans indicate that adr-1 and adr-2 expression peak in early adulthood and decline with age rather precipitously. The declines observed in that study are compatible with a protective role for ADAR alleles in aging.

Prior to this study, RNA editing had not been directly implicated in the regulation of aging in humans or C. elegans. Although the impact upon lifespan that we observed in C. elegans appears to be independent of insulin signaling (Figure 6a), the interaction between RNA editing and RNA interference is likely to be complex since decreased insulin signaling in C. elegans can also affect RNA interference [47] and may suggest threshold effects associated with declining levels of insulin signaling. Future studies will be needed to identify and characterize targets of RNA editing and their potential role(s) in modulating RNA interference activity, in the context of aging and age-related diseases.

Our analysis provides strong evidence for association of ADARB1 and ADARB2 with extreme old age. Our findings of strongest association in the NECS sample are consistent with that sample being both the largest and oldest of the four studies. The lack of reproducibility for some SNPs may have been due to differences in overall genetic background (ethnicity), size and younger ages of the oldest old samples. Nonetheless, associations were noted across four different study populations suggesting that the associations between ADARB1 and ADARB2 and EL are robust to different genetic backgrounds and environmental exposures. ADARB2 is a very large gene spanning more than 500Kb in chromosome 10, but our analysis narrows the association to a region of approximately 100Kb that could be followed-up by fine mapping or sequencing for discovering functional variants and to provide a better understanding of the function of these genes in human aging.

Materials and Methods

Ethic Statement

Subjects included in the NECS, SICS, AJCS and JCS provided written informed consent, and all research involving human subjects was approved by the Institutional Review Boards of Boston University, Boston, USA (NECS), the "Istituto di Ricovero e Cura a Carattere Scientifico "Multimedica, Milano, Italy (SICS), Albert Eistein College of Medicine, Bronx, USA (AJCS), Keio

Table 6. Replication in the JCS subject set.

Gene	Row	SNP	Risk versus referent alleles	CEPH	NECS+SICS OR	NECS+SICS Bayes sig	NECS+SICS p(A)	JPT	JCS (470 oldest old, 538 controls) OR	JCS Bayes sig	JCS p(A)
ADARB2	1	rs10903420	AA v AG/GG	0.327	1.25(1.07;1.45)	0.0020	0.28/0.24	0.058	0.51(0.23;1.07)	0.0397	0.03/0.045
	2	rs1007147	AA v AG/GG	0.312	1.34(1.15;1.59)	0.0003	0.27/0.22	0.091	1.10(0.68;1.81)	0.3435	0.08/0.075
	3	rs2805562	AA v AG/GG	0.15	1.32(1.09;1.59)	0.0022	0.16/0.13	0.058	0.67(0.26;1.56)	0.1813	0.02/0.03
	4	rs884949	AA v AC/CC	0.124	1.24(1.01;1.54)	0.0211	0.12/0.10	0			
	5	rs2805533	AA/AG v GG	0.77	0.86(0.72;1.02)	0.0381	0.79/0.81	0.92	0.59(0.37;0.92)	0.0110	0.88/0.93
	6	rs2387653	AA v AG/GG	0.097	1.21(0.99;1.49)	0.0343	0.15/0.13	0			
	7	rs2805535	AA v AG/GG	0.699	1.37(1.11;1.68)	0.0017	0.74/0.68	0.151	1.02(0.69;1.49)	0.4690	0.14/0.14
	8	rs2805543	AA v AG/GG	0.46	1.24(1.08;1.42)	0.0008	0.53/0.48	0.105	0.86(0.57;1.30)	0.2381	0.10/0.12
	9	rs3898610	AA v AG/GG	0.841	1.36(1.12;1.67)	0.0015	0.87/0.84	0.686	1.09(0.83;1.43)	0.2619	0.66/0.64
	10	rs1533484	AA/AG v GG	0.442	0.86(0.75;0.98)	0.0147	0.38/0.42	0.791	0.79(0.58;1.07)	0.0641	0.76/0.80
	11	rs2676192	AA v AG/GG	0.301	0.83(0.71;0.96)	0.0086	0.26/0.30	0.419	0.92(0.71;1.19)	0.2542	0.40/0.42
	12	rs2387952	AA v AG/GG	0.69	1.20(1.04;1.40)	0.0073	0.72/0.68	0.616	0.98(0.76;1.28)	0.4576	0.59/0.60
	13	rs17294019	AA/AG v GG	0.196	0.78(0.65;0.93)	0.0030	0.17/0.21	0.012	0.54(0.12;2.03)	0.1846	0.01/0.02
ADARB1	14	rs3788157	AA v AG/GG	0.63	1.16(1.00;1.34)	0.0204	0.65/0.62	0.65	1.03(0.79;1.36)	0.4080	0.67/0.66
	15	rs414743	AA/AG v GG	0.47	0.83(0.72;0.96)	0.0053	0.43/0.48	0.31	1.14(0.86;1.52)	0.183	0.29/0.26
	16	rs2838809	AA v AG/GG	0.009	0.46(0.19;1.08)	0.0297	0.005/0.01	0	0.81(0.36;1.81)	0.2959	0.028/0.034
	17	rs2838810	AA/AG v GG	1	2.73(1.11;8.25)	0.0136	0.995/0.99	1	1.14(0.50;2.60)	0.3791	0.974/0.970
	18	rs2838816	AA v AG/GG	0.009	0.48(0.18;1.04)	0.0309	0.005/0.01	0	0.81(0.36;1.79)	0.3032	0.029/0.034

Lists of the SNPs in ADARB2 and ADARB1 that reach statistical significance in the NECS and SICS and were attempted to be replicated in the JCS set. The first 7 columns report the details of the SNPs as described in the legend of Table 3. Columns 8–11 report the results for 14 of the 18 SNPs in Table 3 that were genotyped in 470 oldest old and 538 younger controls of the JCS. Only 3 SNPs reach statistical significance in this set (rows 1, 5 and 10) but one of them have opposite effects compared to the NECS and SICS (row 1). We did not attempt to merge the results from different populations because of the substantial differences in allele frequencies.
doi:10.1371/journal.pone.0008210.t006

University, Tokyo, Japan (JCS). All data were analyzed anonymously. Control data from the Illumina iControlDB database were anonymized.

SNP Genotyping

For the NECS and SICS samples, 1 ug of genomic DNA was analyzed on the Illumina 370 CNV chip (Illumina, San Diego, CA) and only samples with at least a 93% call rate were used for the analysis. For the AJCS and JCS, genotyping was originally performed with the Affymetrix 6.0 chip and 5.0 chips, respectively with required call rates of 99% or greater. Affymetrix Birdseed algorithm and Illumina Beadstudio were used for genotype calling. Non overlapping SNPs that were not approximated by SNPs with substantial LD (D′>0.8) were genotyped with Sequenom (AJCS) and BigDye Terminator cycle sequencing kit and an ABI Prism 3730xl DNA analyzer (Applied Biosystems, CA, USA). The sequence data were analyzed with ABI PRISM SeqScape Software version 2.6 (Applied Biosystems).

Creation of a Genetically Matched Control Set

A referent cohort sample for the NECS subjects was constructed utilizing genotype data from the Illumina iControlDB database and principal components analysis was used to match cases and controls by genetic background. To reduce chances of stratification, we identified 2,077 Caucasian referent subjects from the Illumina iControlDB, all genotype with Illumina arrays, with known age at enrollment between 30 and 75 years, and we used the principal component analysis implemented in the program EIGENSTRAT [48] to examine the structure of this referent group compared to the NECS and SICS subjects. The analysis

showed that both the NECS and the Illumina controls are comprised of three major clusters that correspond to northwest, northeast and southwest Europeans, but in the Illumina controls sample there were also subjects with different levels of admixture between the three clusters Figure S3. We therefore randomly sampled 1,538 subjects from the three major clusters to create a control set that matched the genetic background of the NECS extreme old sample set as suggested in [49]. We use the same procedure to identify 81 female and 16 male subjects to be added to the set of SICS controls. The random-selection procedure was repeated twice and lead to the same results.

Genetic Association Analysis (Pooled DNA Samples)

The statistical analysis of pooling based genome-wide genotype data is described in [27]. Briefly, the method uses Bayesian association tests to score the evidence for allelic associations between centenarians and controls. Prior distributions represent the prior knowledge about the expected number of genes that may be implicated with the trait and therefore correct for multiple comparisons. The analysis also uses linkage disequilibrium (LD) based filters to retain associations that are supported by clusters of SNPs in LD.

Genetic Association Analysis (Individual DNA Samples)

The genotype data of the 31 SNPs in the genes ADARB1 and 114 SNPs in ADARB2 were individually analyzed using Bayesian logistic regression [32] to fit dominant and recessive models of inheritance adjusted by gender. The marginal posterior distributions of the ORs were estimated using the implementation of Gibbs sampling in WinBugs 1.4 [50], and the 2.5th and 97.5th

Table 7. Replication in the JCS subject set using proxy SNPs.

Row	SNP	Chr	position	NECS+SICS			SNP	position	Distance	D'/r2	JPCS (432 oldest old, 346 controls)			Risk versus referent alleles
				OR	Bayes sig	p(A)					OR	Bayes sig	p(A)	
2	rs1007147	10	1341088	1.34(1.15;1.59)	0.0003	0.27/0.22	rs2804099	1346320	5232	0.63/0.3	0.61(0.34;1.07)	0.0422	0.91/0.94	CC/CG v GG
4	rs884949	10	1361610	1.24(1.01;1.54)	0.0211	0.12/0.10	rs10903426	1361386	−224	1.00/0.58	1.70(0.97;3.05)	0.0332	0.09/0.06	CC v CT/TT
6	rs2387653	10	1397826	1.22(0.98;1.48)	0.0343	0.15/0.13	rs17221652	1406472	8646	0.91/0.22	1.84(1.01;3.58)	0.0234	0.08/0.05	CC v CG/GG
7	rs2805535	10	1450432	1.37(1.11;1.68)	0.0017	0.74/0.68	rs4543904	1453158	2726	0.88/0.20	0.44(0.23;0.80)	0.0031	0.91/0.96	CC/CG v GG
18	rs2838816	21	45454470	0.48(0.18;1.04)	0.0309	0.005/0.01	rs6518219	45479145	24675	0.08/0.03	1.36(1.03;1.82)	0.0202	0.50/0.43	AA/AT v TT
							rs2838824	45479813	25343	0.08/0.03	0.54(0.30;0.86)	0.0059	0.89/0.93	CC/CT v TT

Additional SNPs that were genotyped in the JCS set and corroborate associations in the NECS and SICS subjects sets.

percentile were used to estimate 95% credible intervals (CI) for the ORs. The 50th percentile was used to estimate the OR, and the frequency of OR>1 was used to estimate the posterior probability p(OR>1). The Gibbs sampler was run for at least 10,000 iterations and the last 10,000 simulated values were used to estimate these parameters. We used as prior distributions on the regression coefficients of the logit function normal distributions with mean 0 and standard deviation 3.2 that determine a normal prior distribution of the log(OR) with mean 0 (no association) and a variance that ranges between 10 with no genetic effect to 40 with a gene×gender effect. This set of prior distributions was determined to make the analysis robust to rare alleles (frequency<0.10) and we searched for the largest variance that allowed successful execution of the Gibbs sampler. These prior distributions bias the analysis toward the null hypothesis and reduce false positive associations.

The Bayes significance was defined as 1- p(OR>1) when the posterior estimate of the OR was >1, and 1-p(OR<1) when the posterior estimate of the OR was <1, and an association was deemed significant in the NECS, SICS, or the data aggregated from the two studies, if the Bayes significance was smaller than 0.05. This measure of significance is the posterior probability of the null hypothesis $OR \leq 1$ (or $OR \geq 1$) so that small values denotes strong evidence against the null hypothesis [32]. This analysis identified 18 significant SNPs (Table 3), that is more than twice the number expected by chance in 145 independent tests and two SNPs remained significant even after correcting the threshold for the number of tests. Furthermore, the probability that 18 SNPs could be simultaneously found significantly associated under the null hypothesis of no association is 0.0002, using the binomial distribution with n = 145, x = 18 and p = 0.05. An association that was significant in the aggregated NECS and SICS data was deemed replicated in either the AJCS or JCS studies if the same SNP was significant (Bayes significance <0.05) with the same genetic model and consistent effects; or the same SNP did not reach statistical significance (Bayes significance ≥ 0.05), but the ORs in the different studies were in the same direction and when the data from the studies were aggregated, the association was significant. The rationale for the second condition is that both the AJCS and JC have smaller sample sizes, and therefore have less power compared to the NECS. However, consistent effects and increased significance when the aggregated data are analyzed show that the lack of association in the replication study is due to lack of power if effects are similar across different studies. This strategy has been used to increase the power of genetic association studies, see for example [51]. The results are in Tables 4, 5, 6 and 7.

We conducted a similar analysis stratified by gender but the limited sample sizes did not produce strongly significant results.

Linkage Disequilibrium (LD) Heatmaps

We used HaploView 4.1 to create the LD heatmaps and LD displays were generated using the D' color scheme where white represents D' = 0, red represents D' = 1, and different shades of red represent 0<D'<1 (Figures 1 and 4).

Lifespan Measurements in C. elegans

To synchronize worms for lifespan, eggs were isolated (N2, adr-1, adr-2, adr-1;adr-2, rde-1, rde-4, adr-1;adr-2;rde-1) and synchronized by hatching overnight in the absence of food at 20C. Synchronized L1 larvae were counted and plated (10 worms/plate, n = 60) on Escherichia coli bacterial lawns (OP50) on NGM media and allowed to develop to L4-stage larvae at 20C. 5-fluorodeoxyuridine (FudR) solution was added to a final concentration of 0.1mg/
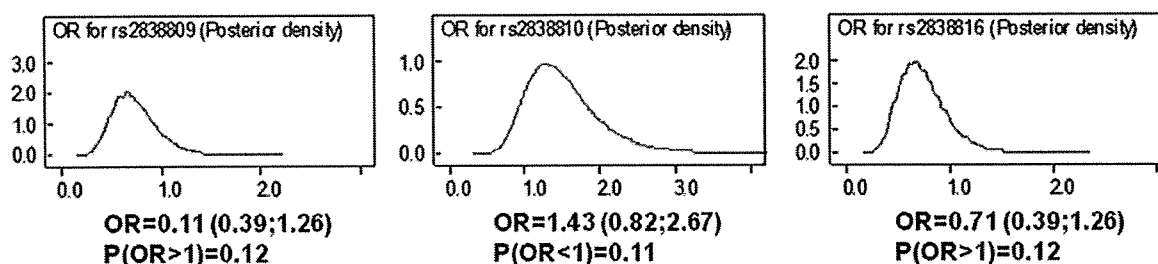
OR=0.11 (0.39;1.26)
P(OR>1)=0.12

OR=1.43 (0.82;2.67)
P(OR<1)=0.11

OR=0.71 (0.39;1.26)
P(OR>1)=0.12

Figure 3. Posterior densities of ORs. Posterior densities of the ORs for the 3 SNPs in ADARB1 with rare alleles and moderate effects in the data aggregated from NECS, SICS and JCS. Significant associations would results in posterior densities not overlapping 1 and definite evidence of either an OR<1 or an OR>1, while all of the three densities have heavy tails and do not provide definite evidence against the null hypothesis of no association.
doi:10.1371/journal.pone.0008210.g003

ml to prevent reproduction. Worms were kept at 20C and lifespan monitored by counting on alternate days. Lifespan was defined as the first day of adulthood (adult lifespan = 0) to death. Aside from reduced lifespan, the worms appeared normal in size and general behavior, consistent with previous reports on *adr* mutant strains [36]. We did observe altered chemotaxis during routine passage of the worms to maintain stocks, as previously noted [36]. We also noted reduced progeny viability (data not shown). However, these
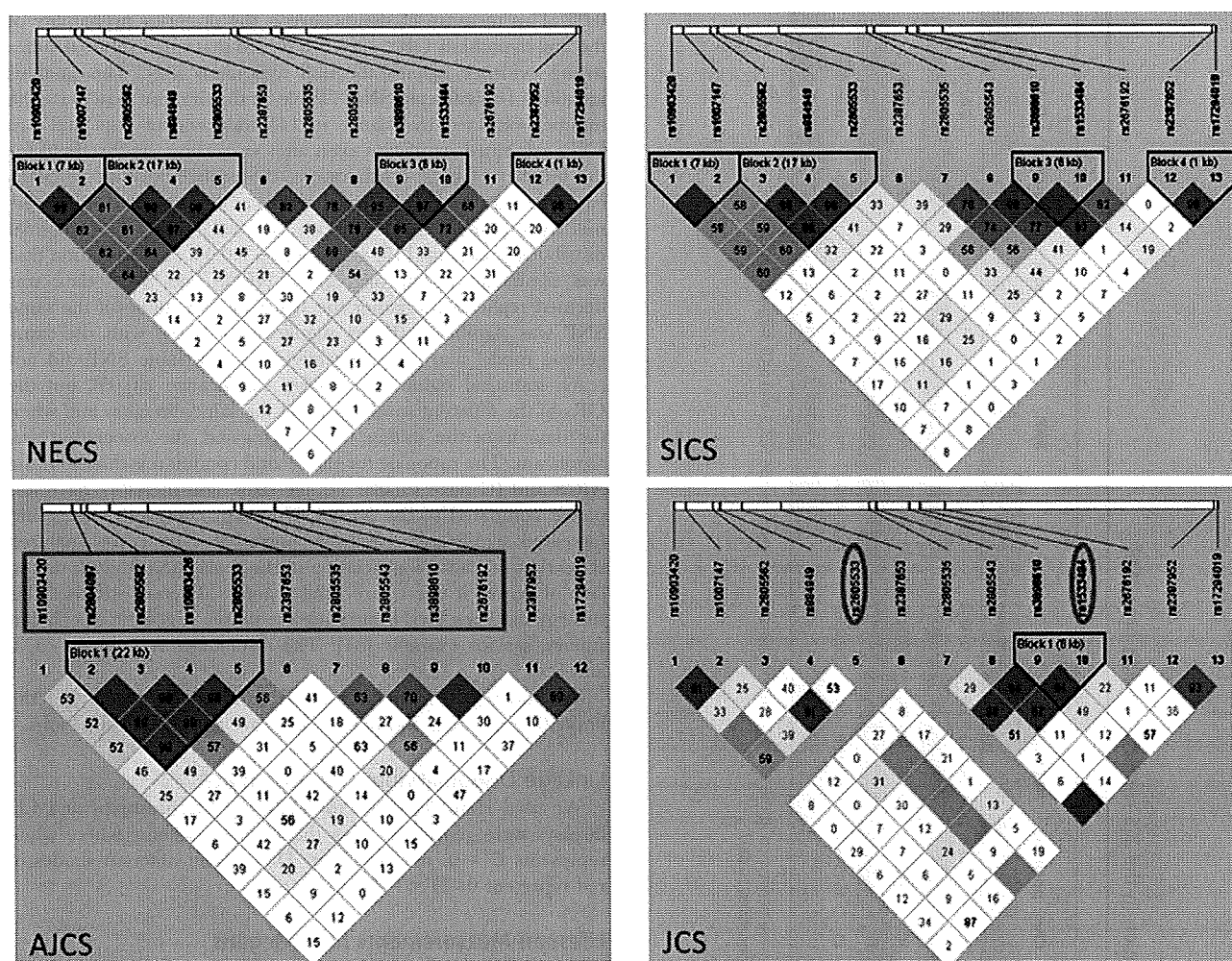

Figure 4. Pattern of LD among the SNP in *ADARB2* (chromosome 10) that are associated with exceptional longevity. The four plots display the pattern of LD captured by the SNPs associated with exceptional longevity in *ADARB2* (chromosome 10) in the NECS, SICS, AJCS and JCS data. The intensity of red cells represents the strength of LD measured by D'. The LD pattern in the NECS, SICS and AJCS subjects are very similar but differ substantially from the pattern of LD in the JCS subjects in which two SNPs become almost monomorphic (rs884949 and rs2387653). Highlighted in red are the SNPs that replicate the results in the AJCS and JCS subjects.
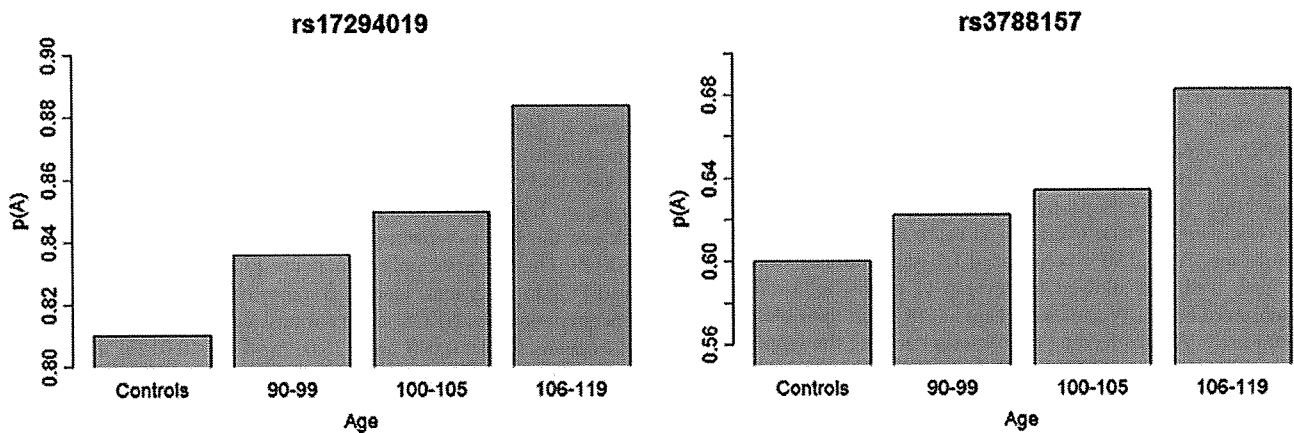doi:10.1371/journal.pone.0008210.g004

rs17294019

rs3788157

Figure 5. Age related trend of allele frequencies. The two barplots show the age related trend of allele frequencies of SNPs rs17294019 (*ADARB2*, SNP # 98 in Table 1) and rs3788157 (ADARB1, SNP # 135 in Table 1) in the NECS (n = 1,023). The frequencies of the common allele for both SNPs were stratified in the age groups 90–99; 100–105, 106 and higher. Trends of allele frequencies for increasing age groups are consistent with a strong correlation between genotype and phenotype that results in substantial enrichment of protective alleles in older subjects.
doi:10.1371/journal.pone.0008210.g005

are unlikely to have influenced our lifespan measurements, since the adult worms were made sterile using FudR and were not transferred during the course of the lifespan assay.

RNAi and Lifespan Measurement

Eggs were isolated from gravid worms and synchronized by hatching overnight in the absence of food. The synchronized L1
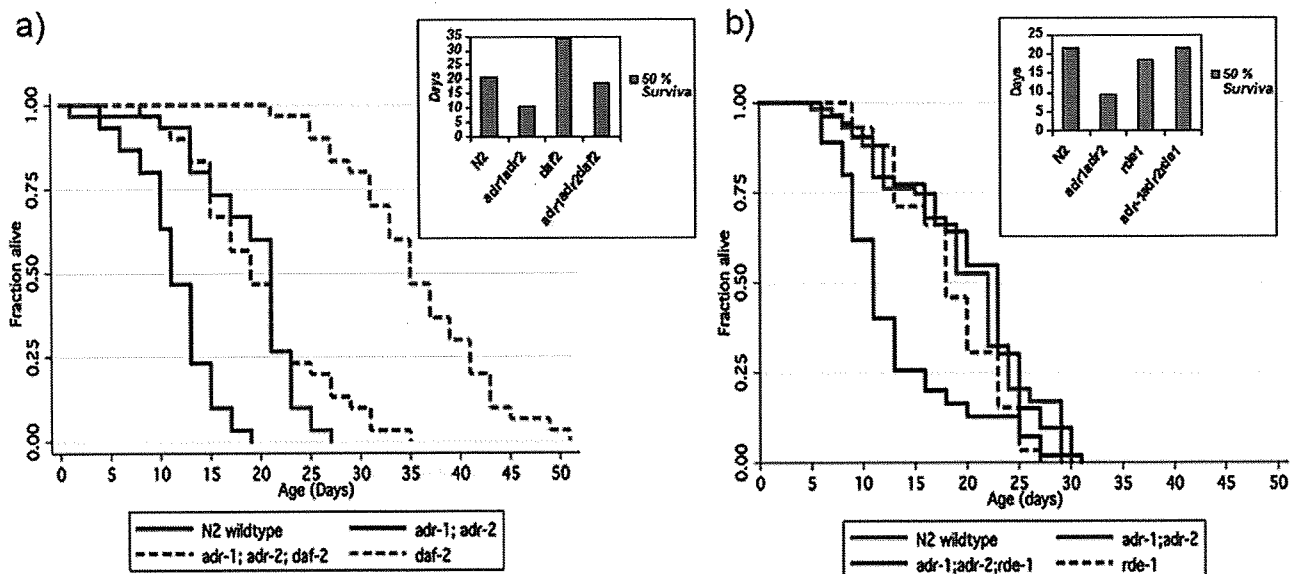

Figure 6. *ADAR* mediated decline in lifespan, *daf-2* influence, and *rde-1* rescue. a) Lifespan using mutant strains for *adr-1;adr-2* in the context of dsRNA mediated gene inactivation of *daf-2*. Synchronized worms at the larval stage 4 (L4) were sterilized with FudR and allowed to feed on bacterial lawns that contained dsRNA for *daf-2*. Note: *adr-1*; *adr-2* double mutant (red solid), *adr-1*; *adr-2* double mutant with dsRNA for *daf-2* (red hatched), N2 wild type (blue solid), N2 with dsRNA for *daf-2* (blue hatched). Note decline in lifespan due to *adr-1*; *adr-2* compared with N2 wildtype. Also note increases in lifespan of both N2 and *adr-1*; *adr-2* in the presence of dsRNA for *daf-2*. The 50% survival time in the *adr-1*; *adr-2* mutant animals was 10 days (95% limits 9 and 12 days) compared with 20 days (95% limits 18 and 20 days) for N2 wild-type control worms. RNAi to *daf-2* increases lifespan to 34 days (95% limits 32 and 40 days), compared with 20 days for the wild type (N2 worms fed empty vector (RNAi)). *daf-2* gene inactivation, in the background of the *adr-1* and *adr-2* null mutations also restored lifespan to 18 days (95% limits 16 and 20 days), compared with 10 days for the *adr-1;adr-2* double mutant strain. b) Lifespan using mutant strains for *adr-1;adr-2* (solid red), N2 wildtype (solid blue), *rde-1* (grey hatched), *adr-1*; *adr-2*; *rde-1* (grey solid) demonstrate declines in lifespan using mutant strains and full rescue of lifespan in an RNAi defective (*rde-1*) background. The *adr-1*; *adr-2* mutant was again about half as long lived as wild-type (median survival time 9 days for *adr-1;adr-2* strain (95% limits 9 and 11 days), and median survival time 21 days (95% limits 18 and 21 days), for N2 wild-type worms. The survival distribution of the triple mutant *adr-1;adr-2; rde-1* is median lifespan 21 days (95% limits 18 and 21 days), which is significantly different from *adr-1;adr-2*, with a median lifespan of 9 days (95% limits 9 and 11 days). The lifespan of *rde-1* was modestly reduced compared with the wild-type N2, as was reported previously[29]. Inset boxes displays 50% survival (days) for each condition and demonstrates that daf-2 gene inactivation increases lifespan, in both wild type and in *adr-1;adr-2* mutant strains (a) and that RNAi knockout (*rde-1*) restores lifespan.
doi:10.1371/journal.pone.0008210.g006

larvae were then placed on OP50-containing agar plates and allowed to develop to L4-stage larvae at 20C. The L4-stage larvae were washed thoroughly, and placed either on *Escherichia coli* HT115 with empty RNAi vector or *Escherichia coli* HT115 expressing double-stranded RNA (dsRNA) for *daf-2*. Briefly, dsRNA -expressing bacteria were grown overnight in LB with 50 ug/ml ampicillin and then seeded onto RNAi NGM plates containing 5 mM isopropylthiogalactoside (IPTG). The RNAi bacteria were induced overnight at room temperature for dsRNA expression. About 30 synchronized L4-stage animals were added to each well and allowed to develop to adults, followed by the addition of FudR. Worms were kept at 20C, and their lifespan was monitored. Worms feeding on bacteria carrying the empty vector were used as a negative control. Log-rank test in the R package survival was used for the statistical analysis.

Supporting Information

Figure S1 Phylogenetic clustering and alignment of ADARs form multiple species. Phylogenetic clustering and alignment of ADARs form multiple species, including Human, chimp, bull, cat, rat, chicken, wolf and nematode (C. elegans). Sequences were aligned by using the neighbor-joining algorithm with Clustal X, gap-stripped with corrections for multiple substitutions, and bootstrap analyzed with 1,000 bootstrap resamplings. Phylograms were generated with NJ-plot. The tree is unrooted and was generated using a Neighbor Joining algorithm implemented in Clustal X. Note bootstrap values defining the three main ADAR branches.
Found at: doi:10.1371/journal.pone.0008210.s001 (0.07 MB TIF)

Figure S2 C. elegans lifespan results using individual gene mutant strains. C. elegans lifespan results using individual gene mutant strains, as indicated. Eggs were isolated from gravid worms and synchronized by hatching overnight in the absence of food. The synchronized L1 larvae were then placed on OP50-containing agar plates and allowed to develop to L4-stage larvae at 20C. The L4-stage larvae were washed thoroughly, and placed on Escherichia coli expressing double-stranded RNA (dsRNA) for daf-2. Briefly, dsRNA -expressing bacteria were grown overnight in LB with 50 ug/ml ampicillin and then seeded onto RNAi NGM plates containing 5 mM isopropylthiogalactoside (IPTG). The RNAi daf-2 bacteria were induced overnight at room temperature for daf-2 dsRNA mediated knockdown expression. About 30 synchronized L4-stage animals were added to each well and allowed to develop to adults, followed by the addition of FudR. Worms were kept at 20C, and their lifespan was monitored. Worms feeding on bacteria carrying the empty vector were used as a negative control. Animals were scored every 1 to 2 days subsequently and scored as dead when they no longer responded to gentle prodding.
Found at: doi:10.1371/journal.pone.0008210.s002 (0.05 MB TIF)

Figure S3 Population structure of NECS centenarians and controls, SICS centenarians and controls, and Illumina controls. Population structure of NECS centenarians (blue) and controls (red), SICS centenarians (green) and controls (orange), and Illumina controls (grey). Each scatter plot shows the first two principal components that were estimated using genotype data for more than 300K SNPs in NECS, SICS and Illumina subjects using the program Eigenstrat. From top to bottom, left to right: (Blue) scatter plot of the first two principal components in centenarians of the NECS. The two principal components (PC1 displayed in the x-axis and PC2 in the y-axis) identify 3 major clusters that based on the ancestry of the NECS centenarians can be labeled as NW Europeans (PC1<0.005 and −0.0125<PC2<0.0125), Ashkenazi Jews (PC1>0.005 and PC2<−0.0125) and SW Europeans/ Italians (PC1>0.005 and PC2>0.0125). The thresholds on the principal components were identified by splitting the components using mixture models. (Green) scatter plot of the first two principal components in centenarians of the SICS. In agreement with the analysis of NECS subjects, the centenarians of the SICS have a SW European genetic background. (Red) scatter plot of the first two principal components in controls of the NECS that display approximately the same population substructure of centenarians; (Grey) scatter plot of the first two principal components in the Illumina controls. The plots show that the controls have a population substructure similar to the NECS cases and controls but also a larger level of admixture between the three European subgroups. (Orange) scatter plot of the first two principal components in controls of the SICS that exhibit the same SW European genetic background of SICS centenarians. Note that the plot of PC1 and PC2 for NECS controls (red) is repeated twice to facilitate the two comparisons within NECS subjects and between NECS and Illumina controls.
Found at: doi:10.1371/journal.pone.0008210.s003 (0.12 MB TIF)

Author Contributions

Conceived and designed the experiments: PS MM MHS GR CB TP. Performed the experiments: PS MM TK MCW EM MM SEJF YA CVA AK GA CB. Analyzed the data: PS MM NS SHH AS AM TP. Contributed reagents/materials/analysis tools: PS MM AAP TK SLA SHH YA AB NB DT AR MS TA YG MHS NH GA GR CB TP. Wrote the paper: PS MM NS MHS GR CB TP.

References

1. Christensen K, Johnson TE, Vaupel JW (2006) The quest for genetic determinants of human longevity: challenges and insights. Nat Rev Genet 7: 436–448.
2. Gudmundsson H, Gudbjartsson DF, Frigge M, Gulcher JR, Stefansson K (2000) Inheritance of human longevity in Iceland. Eur J Hum Genet 8: 743–749.
3. Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, et al. (2006) Genetic influence on human lifespan and longevity. Hum Genet 119: 312–321.
4. Kerber RA, O'Brien E, Smith KR, Cawthon RM (2001) Familial excess longevity in Utah genealogies. J Gerontol A Biol Sci Med Sci 56: B130–139.
5. Perls T, Shea-Drinkwater M, Bowen-Flynn J, Ridge SB, Kang S, et al. (2000) Exceptional familial clustering for extreme longevity in humans. J Am Geriatr Soc 48: 1483–1485.

6. Perls TT, Wilmoth J, Levenson R, Drinkwater M, Cohen M, et al. (2002) Life-long sustained mortality advantage of siblings of centenarians. Proc Natl Acad Sci USA 99: 8442–8447.
7. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, et al. (2006) Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. European Journal of Human Genetics 14: 79–84.
8. Terry DF, Wilcox MA, McCormick MA, Pennington JY, Schoenhofen EA, et al. (2004) Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. J Am Geriatr Soc 52: 2074–2076.
9. Christensen K, McGue M, Petersen I, Jeune B, Vaupel JW (2008) Exceptional longevity does not result in excessive levels of disability. Proc Natl Acad Sci U S A 105: 13274–13279.

10. Hitt R, Young-Xu Y, Silver M, Perls T (1999) Centenarians: the older you get, the healthier you have been. Lancet 354: 652.

11. Terry DF, Sebastiani P, Andersen SL, Perls TT (2008) Disentangling the roles of disability and morbidity in survival to exceptional old age. Arch Intern Med 168: 277–283.

12. Atzmon G, Schechter C, Greiner W, Davidson D, Rennert G, et al. (2004) Clinical phenotype of families with longevity. J Am Geriatr Soc 52: 274–277.

13. Evert J, Lawler E, Bogan H, Perls T (2003) Morbidity profiles of centenarians: survivors, delayers, and escapers. J Gerontol A Biol Sci Med Sci 58: 232–237.

14. Schoenhofen EA, Wyszynski DF, Andersen S, Pennington J, Young R, et al. (2006) Characteristics of 32 supercentenarians. J Am Geriatr Soc 54: 1237–1240.

15. Finch CE, Tanzi RE (1997) Genetics of aging. Science 278: 407–411.

16. Salvioli S, Olivieri F, Marchegiani F, Cardelli M, Santoro A, et al. (2006) Genes, ageing and longevity in humans: problems, advantages and perspectives. Free Radic Res 40: 1303–1323.

17. Sorkin J, Post W, Pollin TI, O'Connell JR, Mitchell BD, et al. (2005) Exploring the genetics of longevity in the Old Order Amish. Mech Ageing Dev 126: 347–350.

18. Perls T (2001) Genetic and phenotypic markers among centenarians. J Gerontol A Biol Sci Med Sci 56: M67–70.

19. Perls T, Kunkel L, Puca A (2002) The genetics of aging. Curr Opin Genet Dev 12: 362–369.

20. Barzilai N, Atzmon G, Schechter C, Schaefer EJ, Cupples AL, et al. (2003) Unique lipoprotein phenotype and genotype associated with exceptional longevity. JAMA 290: 2030–2040.

21. Atzmon G, Rincon M, Schechter CB, Shuldiner AR, Lipton RB, et al. (2006) Lipoprotein genotype and conserved pathway for exceptional longevity in humans. PLoS Biol 4: e113.

22. Atzmon G, Pollin TI, Crandall J, Tanner K, Schechter CB, et al. (2008) Adiponectin levels and genotype: a potential regulator of life span in humans. J Gerontol A Biol Sci Med Sci 63: 447–453.

23. Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, et al. (2008) FOXO3A genotype is strongly associated with human longevity. Proc Natl Acad Sci U S A 105: 13987–13992.

24. Flachsbart F, Caliebe A, Kleindorp R, Blanche H, von Eller-Eberstein H, et al. (2009) Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc Natl Acad Sci U S A 106: 2700–2705.

25. Vijg J, Campisi J (2008) Puzzles, promises and a cure for ageing. Nature 454: 1065–1071.

26. Kojima T, Kamei H, Aizu T, Arai Y, Takayama M, et al. (2004) Association analysis between longevity in the Japanese population and polymorphic variants of genes involved in insulin and insulin-like growth factor 1 signaling pathways. Exp Gerontol 39: 1595–1598.

27. Sebastiani P, Zhao Z, Abad-Grau MM, Riva A, Hartley SW, et al. (2008) A hierarchical and modular approach to the discovery of robust associations in genome-wide association studies from pooled DNA samples. BMC Genetics 9: 6.

28. Yang Q, Rasmussen SA, Friedman JM (2002) Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. Lancet 359: 1019–1025.

29. Yang JH, Luo X, Nie Y, Su Y, Zhao Q, et al. (2003) Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation. Immunology 109: 15–23.

30. Finch CE, Morgan TE (2007) Systemic inflammation, infection, ApoE alleles, and Alzheimer disease: a position paper. Curr Alzheimer Res 4: 185–189.

31. Maas S, Kawahara Y, Tamburro KM, Nishikura K (2006) A-to-I RNA editing and human disease. RNA Biol 3: 1–9.

32. Balding DJ (2006) A tutorial on statistical methods for population association studies. Nat Rev Genet 7: 781–791.

33. Schachter F, Faure-Delanef L, Guenot F, Rouger H, Froguel P, et al. (1994) Genetic associations with human longevity at the APOE and ACE loci. Nat Genet 6: 29–32.

34. Vaupel JW, Carey JR, Christensen K, Johnson TE, Yashin AI, et al. (1998) Biodemographic trajectories of longevity. Science 280: 855–860.

35. Morse DP, Aruscavage PJ, Bass BL (2002) RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. Proc Natl Acad Sci U S A 99: 7906–7911.

36. Tonkin LA, Saccomanno L, Morse DP, Brodigan T, Krause M, et al. (2002) RNA editing by ADARs is important for normal behavior in Caenorhabditis elegans. EMBO J 21: 6025–6035.

37. Keegan LP, Brindle J, Gallo A, Leroy A, Reenan RA, et al. (2005) Tuning of RNA editing by ADAR is required in Drosophila. EMBO J 24: 2183–2193.

38. Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R (1993) A C. elegans mutant that lives twice as long as wild type. Nature 366: 461–464.

39. Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, et al. (2001) Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. Science 292: 104–106.

40. Tatar M, Kopelman A, Epstein D, Tu MP, Yin CM, et al. (2001) A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. Science 292: 107–110.

41. Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloen A, et al. (2003) IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. Nature 421: 182–187.

42. Suh Y, Atzmon G, Cho M-O, Hwang D, Bingrong L, et al. (2008) Functionally significant insulin-like growth factor 1 receptor mutations in centenarians. Proc Natl Acad Sci U S A 105: 3438–3442.

43. Lee SS, Kennedy S, Tolonen AC, Ruvkun G (2003) DAF-16 target genes that control C. elegans life-span and metabolism. Science 300: 644–647.

44. Knight S, Bass B (2002) The role of RNA editing by ADARs in RNAi. Mol Cell Biol 10: 809–817.

45. Tonkin LA, Bass BL (2003) Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. Science 302: 1725.

46. Lund J, Tedesco P, Duke K, Wang J, Kim SK, et al. (2002) Transcriptional profile of aging in C. elegans. Curr Biol 12: 1566–1573.

47. Wang D, Ruvkun G (2004) The insulin pathway regulates RNAi in C. elegans. Cold Spring Harbor Symposium on Quantitative Biology, 69th edition 69: 429–433.

48. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909.

49. Hinds DA, Stokowski RP, Patil N, Konvicka K, Kershenobich D, et al. (2004) Matching strategies for genetic association studies in structured populations. Am J Hum Genet 74: 317–325.

50. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR (1999) Methods in health service research. An introduction to bayesian methods in health technology assessment. BMJ 319: 508–512.

51. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, et al. (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 41: 18–24.

ORIGINAL ARTICLE

Validation of eight genetic risk factors in East Asian populations replicated the association of *BRAP* with coronary artery disease

Kunihiko Hinohara[1,2], Hitoshi Ohtani[1,2], Toshiaki Nakajima[1,2], Taishi Sasaoka[3], Motoji Sawabe[4], Bok-Soo Lee[5], Jimin Ban[5], Jeong-Euy Park[5], Tohru Izumi[3] and Akinori Kimura[1,2]

Coronary artery disease (CAD) is caused by a thrombotic occlusion or spasm of the coronary artery. Association of genetic variants with susceptibility to CAD has been reported in various populations, but the association should be replicated in other populations to establish the role of genetic variants in CAD. We conducted a case–control study with a total of 1480 CAD cases and 2115 controls from two East Asian populations, Japanese and Korean, to validate the associations of CAD with eight single nucleotide polymorphisms (SNPs) in eight loci, which were identified from large-scale whole-genome association studies in Europeans or East Asians. Among the tested SNPs, one SNP in *BRAP* (rs11066001) showed a significant association in allele frequency distribution with CAD in both the Japanese (Odds ratio (OR)=1.63, 95% confidence interval (CI); 1.41–1.89, $P=5.0\times10^{-11}$, corrected P (Pc)$=4.0\times10^{-10}$) and Korean populations (OR=1.68, 95% CI; 1.41–2.00, $P=6.5\times10^{-9}$, $Pc=5.2\times10^{-9}$), and a meta-analysis showed a significant association in the East Asian populations (OR=1.65, 95% CI; 1.48–1.85, $P=1.8\times10^{-18}$, $Pc=1.4\times10^{-17}$), whereas no evidence of association was found for the other SNPs. In addition, a combined analysis of *BRAP* and another CAD locus on 9p21 suggested that these loci had a synergistic role in the susceptibility. Failure to replicate the association with the other SNPs, which were reported in the European populations, suggested that their contributions to CAD were not large enough to be readily captured in the East Asian populations.
Journal of Human Genetics (2009) 54, 642–646; doi:10.1038/jhg.2009.87; published online 28 August 2009

INTRODUCTION

Coronary artery disease (CAD) caused by a thrombotic occlusion or spasm of the coronary artery becomes a major health problem in many countries.[1] CAD is mainly based on coronary atherosclerosis and often manifests with sudden chest pain due to reversible (angina pectoris, AP) or irreversible (myocardial infarction, MI) ischemia in the heart caused by decreased blood flow in the coronary arteries. Although environmental or life-style factors, such as smoking, hypertension, hypercholesterolemia and diabetes mellitus, contribute to the development of CAD,[2] genetic factors are also involved in the pathogenesis of CAD.[3] Several large-scale association studies using a large number of genetic variations, including single nucleotide polymorphisms (SNPs), have recently identified the susceptibility genes and loci for CAD.[4–12] However, not all of the reported associations could be replicated in other studies even if middle- to large-sized samples were investigated in the original reports, suggesting that the contribution of genetic factors was not large enough to be replicated

in some cases.[13] Validation studies for the association in other populations are, therefore, crucial to establish the role of disease-related genes.

Recent progress based on the large-scale association studies has accumulated information on the susceptibility genes linked to CAD. It is noted that seven chromosomal loci were reported to be strongly associated with CAD in Europeans conducted by the Wellcome Trust Case Control Consortium (WTCCC)[6] (loci on chromosomes 1q43, 5q21, 6q25, 9p21, 16q23, 19q12 and 22q12). Samani *et al.*[7] tested several loci in the WTCCC study, which met their predefined criterion ($P<1.2\times10^{-5}$ and less than a 50% chance of being false-positive), in German MI families and found that three loci on chromosomes 2q36, 6q25 and 9p21 were significantly associated with MI. In addition, a combined analysis of data from these two studies revealed four additional CAD-associated loci on chromosomes 1p13, 1q41, 10q11 and 15q22.[7] In a recent study, three other loci were reported to be associated with early-onset MI[10] (loci on chromosomes 2q33, 6q24

[1]Department of Molecular Pathogenesis, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; [2]Laboratory of Genome Diversity, School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan; [3]Department of Internal Medicine and Cardiology, Kitasato University School of Medicine, Sagamihara, Japan; [4]Department of Pathology, Tokyo Metropolitan Geriatric Medical Center, Tokyo, Japan and [5]Division of Cardiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea
Correspondence: Professor A Kimura, Department of Molecular Pathogenesis, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan.
E-mail: akitis@mri.tmd.ac.jp
Received 30 June 2009; revised 10 August 2009; accepted 11 August 2009; published online 28 August 2009

and 21q22). In addition, Gudbjartsson et al.[11] reported that an SNP on chromosome 12q24 was significantly associated with MI. All these studies were conducted in European populations. On the other hand, Ozaki et al.[12] identified the BRCA1-associated protein (BRAP) gene as a possible binding partner of galectin-2.As they reported the gene for galectin-2 LGALS2 as a susceptibility gene for MI,[14] they investigated a possible association of MI with BRAP and found an SNP (rs11066001) to be significantly associated with MI in both the Japanese and Taiwanese populations.[12]

We conducted validation studies in the Japanese and Korean populations, which confirmed the association of CAD with a 9p21 SNP (rs1333049), but failed to replicate the association with SNPs in LTA, LGALS2 and PSMA6, showing that not all of the associations reported in the Japanese population[4,14] could be readily replicated in our studies.[15–17] In this study, we investigated the association between CAD and eight SNPs in eight genome regions in the East Asian populations.

MATERIALS AND METHODS

Subjects
A total of 1480 CAD cases and 2115 control subjects from the Japanese and Korean populations were recruited as reported previously.[15–17] Briefly, the Japanese panels were composed of CAD cases (n=622) and controls (n=1402). The Japanese controls included healthy individuals selected at random (n=656) and healthy-donor-derived B-cell lines obtained from the Japan Health Sciences Foundation (n=746). The Korean subjects consisted of CAD cases (n=858) and controls (n=713). The Korean controls included healthy individuals selected at random (n=179) and cancer patients without CAD (n=534) consisting of 230, 147, 87 and 71 patients with stomach cancer, colon cancer, lung cancer and hepatocellular carcinoma, respectively. The diagnosis of CAD was based on the standard criteria as described previously.[15–17] In brief, the diagnosis of MI was based on typical electrocardiographic changes, increased serum levels of creatinine kinase, aspartate aminotransferase, and lactate dehydrogenase and increased serum concentration of troponin T. The diagnosis was confirmed by the presence of a wall motion abnormality on left ventriculography and by the identification of responsible stenosis in the coronary arteries on coronary angiography. On the other hand, AP was diagnosed by the clinical manifestation of sudden chest pain, transient electrocardiographic changes, and significant stenosis of the coronary arteries found in coronary angiography. Severity of coronary atherosclerosis was classified according to the number of coronary vessels with significant stenosis (angiographic luminal stenosis > 50%) as 0-, 1-, 2- or 3-vessel disease. Informed consent was received from each participant and the study was approved by the Ethics Review Boards of the Medical Research Institute of Tokyo Medical and Dental University, Kitasato University School of Medicine, Tokyo Metropolitan Geriatric Medical Center, and Samsung Medical Center.

Genotyping
Eight SNPs were selected for the validation study, because they were reported to have a minor allele frequency over 0.05 in the Asian populations (Japanese and Han Chinese) in the HapMap data (Supplementary Table 1). All SNPs were genotyped using the TaqMan SNP genotyping assay (Applied Biosystems, Foster City, CA, USA).

Statistical analysis
A statistical analysis for power and sample size computations of case–control design was performed using the PGA program (http://dceg.cancer.gov/bb/tools/pga).[18] Frequencies of alleles or genotypes were compared between the cases and controls using a χ^2-test. The strength of the association was expressed by OR. Meta-analysis was performed using a Mantel–Haenszel method to investigate the association in the combined Japanese and Korean populations. The significance of the association between the severity of coronary atherosclerosis and rs11066001 was examined by using the Mann–Whitney U-test.

A stratification analysis was carried out according to the method described by Svejgaard and Ryder[19] to test the independency or synergistic effects of two

different alleles that were significantly associated with the disease. Briefly, basic data for the analysis were the entries of the two-by-four table giving the four phenotypic combinations of factors A and B in cases and controls, and the data were analyzed in two-by-two tables involving the stratification of each of the two factors against the other. In the analysis, contributions of factors A and B could be evaluated as odds ratios (OR) conferred by one of the factors ($OR_{A, non-B}$ or $OR_{non-A,B}$) or both factors ($OR_{A,B}$) vs non-A non-B. If these factors independently conferred risk of disease, $OR_{A,non-B}$ and $OR_{non-A,B}$ were equal to (or at least not exceeding) $OR_{A,B}$. In contrast, when $OR_{A,B}$ was larger than $OR_{A,non-B}$ or $OR_{non-A,B}$, these factors were considered to synergistically contribute to the risk.

Corrected P (Pc) values were obtained by multiplying the P-values with the number of independent tests (eight for the replication study and eight for the stratification analysis). When the Pc value was less than 0.05, the association was considered to be significant.

RESULTS
Demographic characteristics of the study population are listed in Table 1. This study was conducted to replicate the association of CAD with recently reported genetic risk variants on chromosomes 1p13, 1q41, 1q43, 2q36, 5q21, 10q11 and 16q23. An SNP in BRAP on chromosome 12q24 was also tested for the association. Statistical power analysis of our study design to verify whether it could provide adequate powers in replicating the association with the SNPs on chromosomes 1p13, 1q41, 1q43, 2q36, 5q21, 10q11, 16q23 and 12q24 showed that the statistical powers were 93.3, 99.5, 99.8, 86.1, 99.9, 100.0, 99.2 and 100.0%, respectively, when the Japanese and Korean populations were combined, assuming that they were genetically identical. On the other hand, the powers to detect the association either in the Japanese or Korean populations were 88.6, 98.5, 99.0, 79.7, 99.6, 100.0, 97.7 and 100.0%, respectively, when they were assumed to be genetically different (Supplementary Table 2).

Genotype distributions for all SNPs were in the Hardy–Weinberg equilibrium in the tested populations (data not shown). As shown in Table 2, the SNP in BRAP showed a significant association with CAD in allele distribution in both the Japanese (OR=1.63, 95% CI; 1.41–1.89, $P=5.0\times10^{-11}$, $Pc=4.0\times10^{-10}$) and Korean populations (OR=1.68, 95% CI; 1.41–2.00, $P=6.5\times10^{-9}$, $Pc=5.2\times10^{-8}$). When we performed a meta-analysis of data from the Japanese and Korean populations, the association with CAD was highly significant ($P=1.8\times10^{-18}$, $Pc=1.4\times10^{-17}$) with an OR of 1.65 (95% CI=1.48–1.85) for the risk allele. In addition, when the CAD patients were grouped into AP and MI, the meta-analysis showed that the

Table 1 Clinical characteristics of the study population

	Japanese Population		Korean Population	
	CAD (n=622)	Control (n=1402)	CAD (n=858)	Control (n=713)
Age (years)	59.3±10.2	39.0±10.6	61.2±11.1	58.4±11.5
BMI	23.7±2.9	NA	24.7±2.7	NA
Gender (% male)	83.2	56.3	76.4	64.8
Smoking (%)	73.8	NA	69.3	NA
HT (%)	53.1	NA	45.1	NA
HC (%)	49.9	NA	43.7	NA
DM (%)	30.5	NA	29.1	NA

Characteristics of panels used in this study are shown. The values are means ± s.d. and percentages where indicated. Smoking: current or former smoking of ≥10 cigarettes daily. HT: hypertension, systolic blood pressure of ≥140 mm Hg, diastolic blood pressure of ≥90 mm Hg, or taking antihypertensive medication. HC: hypercholesterolemia, serum total cholesterol of ≥5.72 mmol l⁻¹ (220 mg per 100 ml) or taking lipid-lowering medication. DM: diabetes mellitus, fasting blood glucose of ≥6.93 mmol l⁻¹ (126 mg per 100 ml), glycosylated hemoglobin of ≥6.5%, or taking anti-diabetes medication. NA, data not available.

Table 2 Validation of eight genetic risk variants for the association with CAD

Chromosome or gene	SNP	Minor Allele	Risk Allele	Study Population	Control Number	Case Number	Control MAF	Case MAF	OR (95% CI) for risk allele	P-value	Pc
1p13	rs599839	G	A	Japanese	1384	551	0.068	0.054	1.28 (0.95–1.73)	0.11	NS
				Korean	674	704	0.070	0.075	0.94 (0.71–1.26)	0.68	NS
				Meta-analysis	—	—	0.069	0.065	1.09 (0.89–1.35)	0.39	NS
1q41	rs17465637	A	C	Japanese	1343	575	0.471	0.449	1.09 (0.95–1.25)	0.21	NS
				Korean	641	760	0.440	0.418	1.09 (0.94–1.27)	0.24	NS
				Meta-analysis	—	—	0.461	0.431	1.09 (0.99–1.21)	0.086	NS
1q43	rs17672135	C	T	Japanese	1321	620	0.169	0.186	0.89 (0.75–1.06)	0.19	NS
				Korean	675	740	0.119	0.109	1.10 (0.87–1.39)	0.41	NS
				Meta-analysis	—	—	0.154	0.145	0.96 (0.84–1.11)	0.59	NS
2q36	rs2943634	A	C	Japanese	1373	553	0.102	0.092	1.12 (0.88–1.42)	0.36	NS
				Korean	636	669	0.087	0.090	0.96 (0.73–1.26)	0.78	NS
				Meta-analysis	—	—	0.097	0.091	1.05 (0.88–1.25)	0.61	NS
5q21	rs383830	A	T	Japanese	1392	596	0.188	0.180	1.06 (0.89–1.26)	0.52	NS
				Korean	707	775	0.208	0.225	0.91 (0.76–1.08)	0.27	NS
				Meta-analysis	—	—	0.195	0.205	0.98 (0.87–1.11)	0.75	NS
10q11	rs501120	C	T	Japanese	1387	565	0.335	0.310	1.12 (0.97–1.30)	0.13	NS
				Korean	688	728	0.384	0.361	1.10 (0.95–1.28)	0.22	NS
				Meta-analysis	—	—	0.351	0.339	1.11 (1.00–1.24)	0.051	NS
16q23	rs8055236	T	G	Japanese	1352	620	0.140	0.135	1.04 (0.86–1.26)	0.69	NS
				Korean	667	720	0.096	0.133	0.69 (0.55–0.88)	0.0025	0.02
				Meta-analysis	—	—	0.126	0.134	0.88 (0.76–1.03)	0.10	NS
BRAP (12q24)	rs11066001	G	G	Japanese	1402	603	0.243	0.344	1.63 (1.41–1.89)	5.0×10^{-11}	4.0×10^{-10}
				Korean	713	757	0.182	0.272	1.68 (1.41–2.00)	6.5×10^{-9}	5.2×10^{-8}
				Meta-analysis	—	—	0.223	0.304	1.65 (1.48–1.85)	1.8×10^{-18}	1.4×10^{-17}

Abbreviations: CAD, coronary artery disease; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; Pc, P-value after Bonferroni's correction; NS, not significant (P>0.05).
For the meta-analysis of combined Japanese and Korean groups, OR and P-value were calculated using a Mantel–Haenszel model. Numbers of controls and cases were not indicated for the meta-analysis because they were indicated separately in the Japanese and Korean populations.

association was significant in both groups; OR=1.49, 95% CI=1.22–1.82, $Pc=7.0 \times 10^{-5}$ for AP ($n=369$) and OR=1.71, 95% CI=1.52–1.93, $Pc=1.4 \times 10^{-19}$ for MI ($n=1360$). To investigate the mode of inheritance, we calculated genotype-specific ORs for rs11066001 in BRAP, and the results showed that ORs for AA, AG and GG genotypes were 0.55, 1.52 and 2.04, respectively, suggesting the additive model (Supplementary Table 3).

The association with the SNP in BRAP was investigated in relation to the presence of classical risk factors including gender, age, smoking, hypertension, hypercholesterolemia, obesity and diabetes mellitus, and the severity of coronary atherosclerosis (number of vessels with significant luminal stenosis) in patients whose coronary angiographic data were available (598 Japanese cases and 453 Korean cases). It was observed that the association was not depending on the classical risk factors, suggesting that the BRAP SNP was an independent risk factor for CAD (data not shown). On the other hand, the G allele frequencies of rs11066001 in the Japanese CAD patients with 0, 1, 2 and 3 vessel disease were 0.36 (vs 0.24 in the Japanese controls, $n=18$, OR=1.76, Pc=ns), 0.32 ($n=247$, OR=1.49, $Pc=1.2 \times 10^{-3}$), 0.34 ($n=163$, OR=1.63, $Pc=6.8 \times 10^{-4}$), and 0.37 ($n=170$, OR=1.83, $Pc=3.1 \times 10^{-6}$), respectively, showing that the association was relatively strong in patients with severe CAD but the trend was not statistically significant (Mann–Whitney's U-test, P=ns). In contrast, the data in the Korean CAD patients were 0.18 (vs 0.18 in the Korean controls, $n=11$, OR=1.00, Pc=ns), 0.26 ($n=210$, OR=1.58, $Pc=3.6 \times 10^{-3}$), 0.34 ($n=148$, OR=2.33, $Pc=6.4 \times 10^{-9}$) and 0.35 ($n=84$, OR=2.38, $Pc=4.0 \times 10^{-6}$), respectively, showing a significant correlation between the G-allele frequency and the severity of coronary atherosclerosis (Mann–Whitney's U-test, $P=5.0 \times 10^{-3}$). On the other

hand, no significant association for East Asians was found with the other seven SNPs (Table 2), even when the patients were stratified by gender, onset age and the presence of classical risk factors (data not shown).

As we have previously reported the association of 9p21 SNP with CAD in the Japanese and Korean populations,[15] it was interesting to know whether there was an interaction between 9p21 and BRAP SNP in susceptibility to CAD. We stratified the cases and controls that were genotyped for both 9p21 and BRAP SNPs by the presence or absence of risk genotypes, GG genotype of rs11066001 (BRAP) and CC genotype of rs1333049 (9p21) (Table 3). Separate analyses in the Japanese and Korean populations did not give conclusive results when the P-values were corrected for multiple tests, presumably because the numbers of samples were not large enough. However, the meta-analysis revealed significant associations for BRAP (test {1}, $OR_A=2.01$, $Pc=1.3 \times 10^{-5}$) and 9p21 (test {2}, $OR_B=1.42$, $Pc=9.6 \times 10^{-4}$) and the BRAP genotype increased the risk in both the presence and absence of the 9p21 genotype (test {3}, OR=2.31, Pc=0.018 and test {4}, $OR_{A,non-B}=1.90$, $Pc=1.3 \times 10^{-3}$, respectively). In addition, contribution of the 9p21 genotype was significant in the absence of BRAP genotype (test {6}, $OR_{non-A,B}=1.39$, $Pc=3.4 \times 10^{-3}$). The OR in the presence of both the genotypes (test {8}, $OR_{A,B}=3.33$, $Pc=2.4 \times 10^{-4}$) was higher than those in the presence of one genotype (tests {4} and {6}), implying a synergistic contribution of these two genetic risk factors.

DISCUSSION

Recent genome-wide association studies and candidate gene approaches have identified several susceptibility loci and genes for CAD.