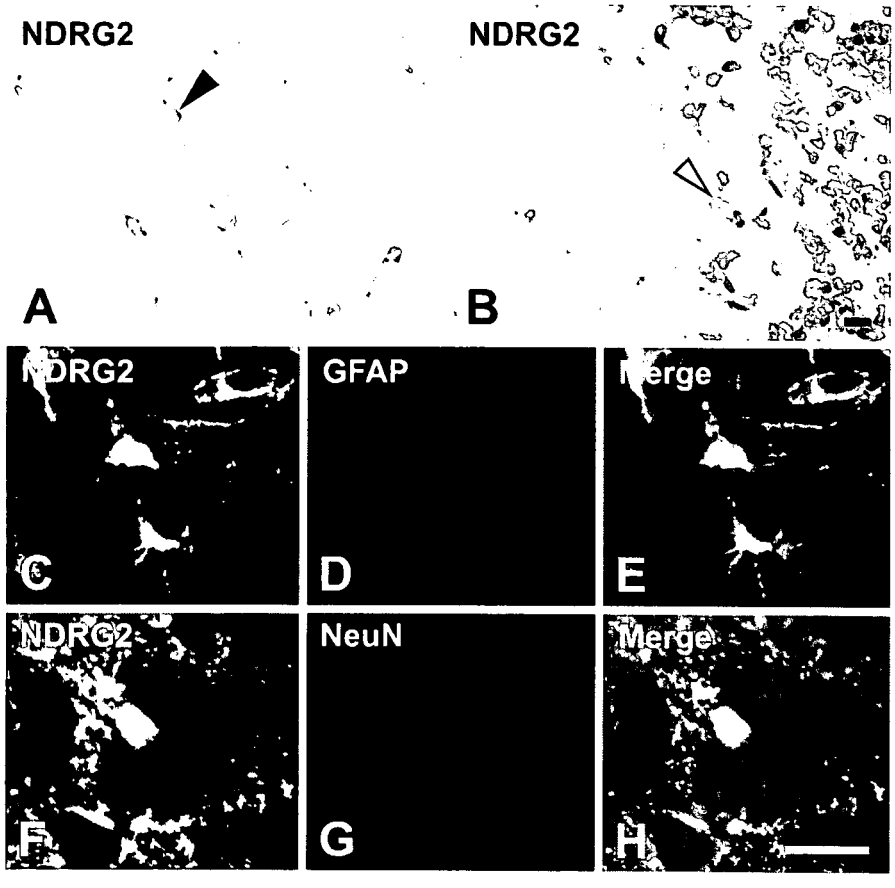

Figure 5 Localization of NDRG2 in the brain. NDRG2 was detected in the astrocytes in the cerebrum (arrowhead in A) and in Bergmann glial cells in the cerebellum (open arrowhead in B). Expression of NDRG2 (C,F) was colocalized with an astrocyte-specific marker GFAP (D), but not with a neuron marker NeuN (G). Merged images are shown in E and H, respectively. Bar = 10 μm.

the expression patterns of NDRGs in the mouse CNS. The expression patterns of each NDRG are summarized in Table 1.

Antibody specificity was examined by Western blotting analysis using recombinant NDRG–GFP fusion proteins. Each antibody specifically reacted with the corresponding protein, although the anti-NDRG1 antibody exhibited a weak cross-reaction to NDRG3 (Figure 1). This may be caused by a relatively high identity of NDRG3 to NDRG1 in amino acid sequence compared

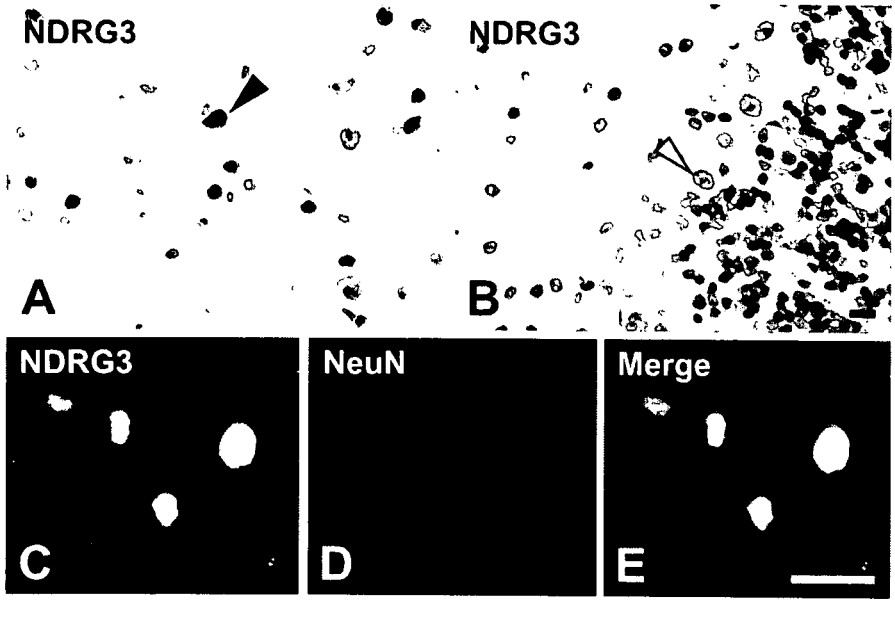

Figure 6 Localization of NDRG3 in the brain. NDRG3 was detected in the nucleus of most cells in the cerebrum, but especially strongly in the neurons (arrowhead in A) and nucleus of Purkinje cells in the cerebellum (open arrowhead in B). Expression of NDRG3 (C) was colocalized with a neuron marker NeuN (D). Merged images are shown in E. Bar = 10 μm.

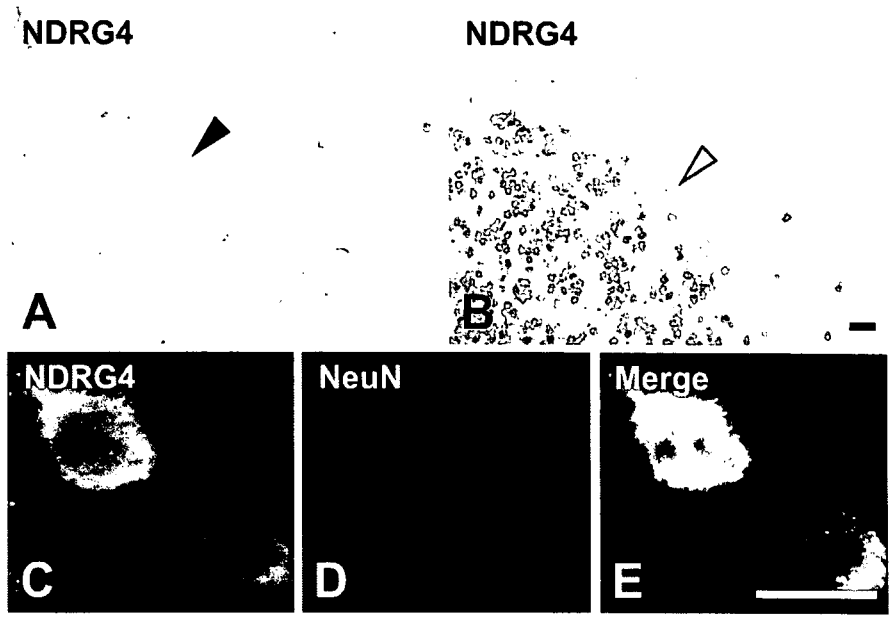

Figure 7 Localization of NDRG4 in the brain. NDRG4 was detected in most cells in the cerebrum, but especially strongly in the neurons (arrowhead in **A**) and Purkinje cells in the cerebellum (open arrowhead in **B**). Strong expression of NDRG4 (**C**) was colocalized with a neuron marker NeuN (**D**). Merged images are shown in **E**. Bar = 10 μm.

with NDRG2 and NDRG4. Immunohistochemical analysis, however, could fortunately discriminate NDRG1 from NDRG3: anti-NDRG1 showed a cytoplasmic staining pattern in particular cells, whereas anti-NDRG3 reacted with the nuclei in most cells of the brain.

We demonstrated here that NDRG1 was mainly localized in the oligodendrocytes (Figure 4). Another group (Wakisaka et al. 2003), however, has reported immunohistochemical data inconsistent with the present study. That report demonstrated that the localization of NDRG1 is changed from hippocampal neurons to astrocytes during postnatal development in the rat brain. Although the inconsistency may be caused by the difference in animal species or developmental process, the possibility of unexpected cross-reactions of their antibody to other NDRGs (probably NDRG2) cannot be ruled out. In fact, our observation of NDRG1 localization in oligodendrocytes is consistent with that of another report (Berger et al. 2004).

The oligodendrocyte is a glial cell engaged in the formation of myelin sheaths in the CNS, whereas the Schwann cell expressing NDRG1 plays an analogous role in the PNS. NDRG1, therefore, may contribute to cellular processes in the development or maintenance of myelin sheaths. Although the loss of NDRG1 in Schwann cells led to demyelination in the sciatic nerves (Okuda et al. 2004), the loss in oligodendrocytes had no effect in the brain (Figure 3). These observations suggested that other NDRGs may compensate for the NDRG1 deficiency in oligodendrocytes but cannot do so in Schwann cells. In fact, all NDRGs except NDRG1 were less expressed in the sciatic nerve than in the brain (Okuda et al. 2004).

NDRG2 was localized to the astrocytes in the cerebrum and to Bergmann glial cells in the cerebellum (Figure 5). NDRG3 was expressed in most cells in the cerebrum and cerebellum, and the subcellular localization of NDRG3 was restricted in the nucleus (Figure 6). These marked differences from NDRG1 in the cellular and subcellular localization suggested that NDRG2 and NDRG3 may not have a redundant function of NDRG1. In fact, NDRG2 and NDRG3 were unable to compensate for the NDRG1 deficiency in sciatic nerves despite their expression in the tissue (Okuda et al. 2004).

In contrast to NDRG2 and NDRG3, NDRG4 may be a likely candidate of compensators for the NDRG1 deficiency in the brain. NDRG4 was abundantly expressed in the brain, especially in the neurons and Purkinje cells (Figure 7), the latter of which were also rich in NDRG1. NDRG1 was originally identified as a gene upregulated with homocysteine treatment (Kokame et al. 1996), and expression of NDRG4 is also induced by homocysteine (Nishimoto et al. 2003). These similarities between NDRG1 and NDRG4 may signify their functional similarity. Failure in compensation for the loss of NDRG1 in the *Ndrg1*-deficient PNS can be explained by the fact that there is little expression of NDRG4 in the sciatic nerves (Okuda et al.

Table 1 Summary of major expression cells of NDRG family proteins in the brain

	NDRG1	NDRG2	NDRG3	NDRG4
Cerebrum	Oligodendrocytes Ependymal cells	Astrocytes	Most cells (nucleus)	Most cells
Cerebellum	Purkinje cells	Bergmann glia	Purkinje cells (nucleus) Most cells (nucleus)	Purkinje cells

2004). Further analysis, however, will be needed to clarify the functional specificity and redundancy of NDRGs in the CNS and also in other physiological systems. Developing and analyzing knockout mice for NDRG2, NDRG3, and NDRG4 would be the most effective approach.

Acknowledgments

This work was supported in part by Grants-in-Aid from the Ministry of Health, Labor, and Welfare of Japan; the Ministry of Education, Culture, Sports, Science, and Technology of Japan; and the Program for the Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NIBIO).

Literature Cited

Agarwala KL, Kokame K, Kato H, Miyata T (2000) Phosphorylation of RTP, an ER stress-responsive cytoplasmic protein. Biochem Biophys Res Commun 272:641–647

Bandyopadhyay S, Pai SK, Gross SC, Hirota S, Hosobe S, Miura K, Saito K, et al. (2003) The Drg-1 gene suppresses tumor metastasis in prostate cancer. Cancer Res 63:1731–1736

Berger P, Sirkowski EE, Scherer SS, Suter U (2004) Expression analysis of the N-Myc downstream-regulated gene 1 indicates that myelinating Schwann cells are the primary disease target in hereditary motor and sensory neuropathy-Lom. Neurobiol Dis 17:290–299

Guan RJ, Ford HL, Fu Y, Li Y, Shaw LM, Pardee AB (2000) Drg-1 as a differentiation-related, putative metastatic suppressor gene in human colon cancer. Cancer Res 60:749–755

Kalaydjieva L, Gresham D, Gooding R, Heather L, Baas F, de Jonge R, Blechschmidt K, et al. (2000) N-myc downstream-regulated gene 1 is mutated in hereditary motor and sensory neuropathy-Lom. Am J Hum Genet 67:47–58

Kalaydjieva L, Hallmayer J, Chandler D, Savov A, Nikolova A, Angelicheva D, King RH, et al. (1996) Gene mapping in Gypsies identifies a novel demyelinating neuropathy on chromosome 8q24. Nat Genet 14:214–217

Kokame K, Kato H, Miyata T (1996) Homocysteine-respondent genes in vascular endothelial cells identified by differential display analysis. GRP78/BiP and novel genes. J Biol Chem 271: 29659–29665

Kovacevic Z, Richardson DR (2006) The metastasis suppressor, Ndrg-1: a new ally in the fight against cancer. Carcinogenesis 27: 2355–2366

Kurdistani SK, Arizti P, Reimer CL, Sugrue MM, Aaronson SA, Lee SW (1998) Inhibition of tumor cell growth by RTP/rit42 and its responsiveness to p53 and DNA damage. Cancer Res 58: 4439–4444

Nishimoto S, Tawara J, Toyoda H, Kitamura K, Komurasaki T (2003) A novel homocysteine-responsive gene, smap8, modulates mitogenesis in rat vascular smooth muscle cells. Eur J Biochem 270:2521–2531

Okuda T, Higashi Y, Kokame K, Tanaka C, Kondoh H, Miyata T (2004) Ndrg1-deficient mice exhibit a progressive demyelinating disorder of peripheral nerves. Mol Cell Biol 24:3949–3956

Okuda T, Kondoh H (1999) Identification of new genes Ndr2 and Ndr3 which are related to Ndr1/RTP/Drg1 but show distinct tissue specificity and response to N-myc. Biochem Biophys Res Commun 266:208–215

Piquemal D, Joulia D, Balaguer P, Basset A, Marti J, Commes T (1999) Differential expression of the RTP/Drg1/Ndr1 gene product in proliferating and growth arrested cells. Biochim Biophys Acta 1450:364–373

Qu X, Zhai Y, Wei H, Zhang C, Xing G, Yu Y, He F (2002) Characterization and expression of three novel differentiation-related genes belong to the human NDRG gene family. Mol Cell Biochem 229:35–44

Shimono A, Okuda T, Kondoh H (1999) N-myc-dependent repression of *Ndr1*, a gene identified by direct subtraction of whole mouse embryo cDNAs between wild type and *N-myc* mutant. Mech Dev 83:39–52

Stein S, Thomas EK, Herzog B, Westfall MD, Rocheleau JV, Jackson RS 2nd, Wang M, et al. (2004) NDRG1 is necessary for p53-dependent apoptosis. J Biol Chem 279:48930–48940

van Belzen N, Dinjens WN, Diesveld MP, Groen NA, van der Made AC, Nozawa Y, Vlietstra R, et al. (1997) A novel gene which is up-regulated during colon epithelial cell differentiation and down-regulated in colorectal neoplasms. Lab Invest 77:85–92

Wakisaka Y, Furuta A, Masuda K, Morikawa W, Kuwano M, Iwaki T (2003) Cellular distribution of NDRG1 protein in the rat kidney and brain during normal postnatal development. J Histochem Cytochem 51:1515–1525

Xu B, Lin L, Rote NS (1999) Identification of a stress-induced protein during human trophoblast differentiation by differential display analysis. Biol Reprod 61:681–686

Zhou D, Salnikow K, Costa M (1998) *Cap43*, a novel gene specifically induced by Ni^{2+} compounds. Cancer Res 58:2182–2189

Zhou RH, Kokame K, Tsukamoto Y, Yutani C, Kato H, Miyata T (2001) Characterization of the human NDRG gene family: a newly identified member, NDRG4, is specifically expressed in brain and heart. Genomics 73:86–97

Human Molecular Genetics, 2008, Vol. 17, No. 4 **617–627**
doi:10.1093/hmg/ddm335
Advance Access published on November 14, 2007

High-density association study and nomination of susceptibility genes for hypertension in the Japanese National Project

Norihiro Kato[1,*], Toshiyuki Miyata[3], Yasuharu Tabara[4], Tomohiro Katsuya[6], Kazuyuki Yanai[1], Hironori Hanada[3], Kei Kamide[3], Jun Nakura[5], Katsuhiko Kohara[5], Fumihiko Takeuchi[2], Hiroyuki Mano[7], Michio Yasunami[8,9], Akinori Kimura[8,9], Yoshikuni Kita[10], Hirotsugu Ueshima[10], Tomohiro Nakayama[11], Masayoshi Soma[12], Akira Hata[13], Akihiro Fujioka[14], Yuhei Kawano[3], Kazuwa Nakao[15], Akihiro Sekine[16], Teruhiko Yoshida[17], Yusuke Nakamura[16,18], Takao Saruta[19], Toshio Ogihara[6], Sumio Sugano[20], Tetsuro Miki[5] and Hitonobu Tomoike[3]

[1]Department of Gene Diagnostics and Therapeutics and [2]Department of Medical Ecology and Informatics, Research Institute, International Medical Center of Japan, Tokyo, Japan, [3]National Cardiovascular Center, Suita, Osaka, Japan, [4]Department of Basic Medical Research and Education and [5]Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Toon, Ehime, Japan, [6]Department of Geriatric Medicine, Osaka University Graduate School of Medicine, Suita, Osaka, Japan, [7]Division of Functional Genomics, Jichi Medical University, Shimotsuke-shi, Tochigi, Japan, [8]Department of Molecular Pathogenesis, Medical Research Institute and [9]Laboratory of Genome Diversity, School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan, [10]Department of Health Science, Shiga University of Medical Science, Otsu, Shiga, Japan, [11]Division of Molecular Diagnostics, Advanced Medical Research Center and [12]Division of Nephrology and Endocrinology, Department of Internal Medicine, Nihon University School of Medicine, Tokyo, Japan, [13]Department of Public Health, Graduate School of Medicine, Chiba University, Chiba, Japan, [14]Amagasaki Health Medical Foundation, Amagasaki, Hyogo, Japan, [15]Department of Medicine and Clinical Science, Kyoto University Graduate School of Medicine, Kyoto, Japan, [16]SNP Research Center, Institute of Physical and Chemical Research, Tokyo, Japan, [17]Genetics Division, National Cancer Center Research Institute, Tokyo, Japan, [18]Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan, [19]Department of Internal Medicine, School of Medicine, Keio University, Tokyo, Japan and [20]Division of Bioscience, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan

Received August 23, 2007; Revised and Accepted November 12, 2007

Essential hypertension is one of the most common, complex diseases, of which considerable efforts have been made to unravel the pathophysiological mechanisms. Over the last decade, multiple genome-wide linkage analyses have been conducted using 300–900 microsatellite markers but no single study has yielded definitive evidence for 'principal' hypertension susceptibility gene(s). Here, we performed a three-tiered, high-density association study of hypertension, which has been recently made possible. For tier 1, we genotyped 80 795 SNPs distributed throughout the genome in 188 male hypertensive subjects and two general population control groups (752 subjects per group). For tier 2 (752 hypertensive and 752 normotensive subjects), we genotyped a panel of 2676 SNPs selected (odds ratio $\geq$ 1.4 and $P \leq 0.015$ in tier 1) and identified 75 SNPs that showed similar tendency of association in tier 1 and tier 2 samples ($P \leq 0.05$ for allele frequency and $P \leq 0.01$ for genotype distribution tests). For tier 3 (619 hypertensive and 1406 normotensive subjects), we genotyped the 75 SNPs and found nine SNPs from seven genomic loci to be associated with hypertension

*To whom correspondence should be addressed at: Department of Gene Diagnostics and Therapeutics, Research Institute International Medical Center of Japan, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan. Tel: +81 332027181; Fax: +81 332027364; Email: nokato@ri.imcj.go.jp

($P \leq 0.05$). In three of these loci, the lowest *P*-values were observed for rs3755351 ($P = 1.7 \times 10^{-5}$) in *ADD2*, rs3794260 ($P = 0.0001$) in *KIAA0789* and rs1805762 ($P = 0.0003$) in *M6PR* when case–control comparison was made in the combined data. An SNP (rs3755351) within *ADD2* had the lowest *P*-value and its experiment-wide significance level is 0.13. Thus, these results have nominated several susceptibility genes for hypertension, and independent replication will clarify their etiological relevance.

INTRODUCTION

Essential hypertension (MIM 145500) is a multifactorial trait, in which interactions among genetic, environmental and demographic factors are involved. Substantial contribution of genetic factors to the overall disease etiology has been documented by a number of epidemiological studies. For example, family studies controlling for a common environment indicate that blood pressure heritability is in the range of 15–35% (1–3). Accordingly, considerable efforts have been made in the study of molecular genetics of hypertension, but the inherently complex nature has hampered progress in the elucidation of the genes involved (4). Over the last decade, multiple genome-wide linkage analyses have been conducted by using microsatellite markers to localize genes influencing hypertension status and/or blood pressure levels in a number of populations derived from various ethnic groups. Although no single study has so far yielded definitive evidence for 'principal' hypertension susceptibility gene(s), some of these studies provide consistency of linkage results in a few chromosomal regions (5–7). It is therefore assumed that multiple genes contribute to the etiology of hypertension independently or synergistically, with each gene exerting small effects under a certain environmental condition.

In parallel with family-based linkage analyses across the entire genome, population-based association studies have been performed, particularly focusing on individual candidate genes to search for genetic influences on hypertension. Association studies for mapping disease-related genes have recently gained popularity over traditional family-based linkage analyses mainly because of their far greater statistical power to detect the presence of genes with relatively 'minor' effects (8,9). Some researchers criticize the liability to false-positive or non-replicable claims. Nevertheless, population-based association studies have become an alternative and complementary approach to family-based linkage analyses in practice.

Given the limitation of statistical power that can be achieved by family-based linkage analyses with sample size practically collectable, population-based association studies are now underway in a genome-wide scale for a number of multifactorial diseases (10). Here, we performed a high-density association study of hypertension with a three-tiered genotyping approach in the Japanese population (Fig. 1).

RESULTS

Multi-tiered case–control study

We performed a large-scale case–control association study of hypertension using SNP markers selected from the Japanese SNP (JSNP) database (11,12). These SNP markers were distributed throughout the genome (Table 1). Only male hypertensive individuals were tested in tier 1, and a total of 80 795 SNPs distributed on 22 autosomes were used for the association study. Details of the high-throughput genotyping were same as previously described (13,14), and technical evaluation of our genotyping assay (e.g. overall success rate and accuracy of the genotyping assay) is shown in the supplementary material (Supplementary Explanation). JSNP had been developed as a database for the SNP discovery project with particular focus on common gene variations in the Japanese population. Although SNP marker resources used in the current study showed a certain degree of diversity in terms of the number of typed SNPs per gene locus, this partially reflected the variable size of re-sequenced fragments depending on the individual gene structure (12).

The gene-centered genome-wide exploratory test in tier 1 identified 2676 SNPs with odds ratio (OR) ≥ 1.4 and $P \leq 0.015$ in at least one test comparing allele frequency and/or genotype distribution (dominant or recessive models) between 188 hypertensive patients and 752 population control subjects in either of two panels (see Materials and Methods). In this exploratory test, the SNPs showing inverted tendency of OR between two pairs of case–control comparisons and significant deviations from Hardy–Weinberg equilibrium (HWE) in any panel ($P \leq 0.01$) were excluded. Subsequently, we performed a screening of these 2676 SNPs with 752 hypertensive patients and 752 normotensive controls in tier 2, which constituted the first 'case versus unaffected control' study panel, i.e. comparison between 940 cases and 752 controls, together with the 188 cases in tier 1. On the basis of relatively stringent criteria, we identified 75 SNPs that showed *P*-values of ≤ 0.01 for genotype distribution and *P*-values of ≤ 0.05 for allele frequency in the χ^2-test statistic. To further examine the association signals, we performed a replication study of these 75 SNPs with another panel of 619 hypertensive subjects and 1406 normotensive controls in tier 3. Cases and unaffected controls collected in tiers 2 and 3 were enrolled according to the identical criteria and their baseline characteristics are shown in Table 2. There were some trait differences in cases between tiers 2 and 3, such as blood pressure measurements and percentages of the subjects taking anti-hypertensive medication. This could be largely attributed to differences in sample enrollment settings between tiers 2 and 3; that is, cases in tier 3 were enrolled from either the annual medical checkup of a medical institution or the clinic practices of general practitioners, whereas a major part of cases in tier 2 were from the clinic practices of university hospitals. Among the 75 SNPs showing *P*-values between 0.05 and 4.4×10^{-5} in the first 'case versus unaffected control' study, only nine SNPs showed borderline association (at the level of $P \leq 0.05$) in

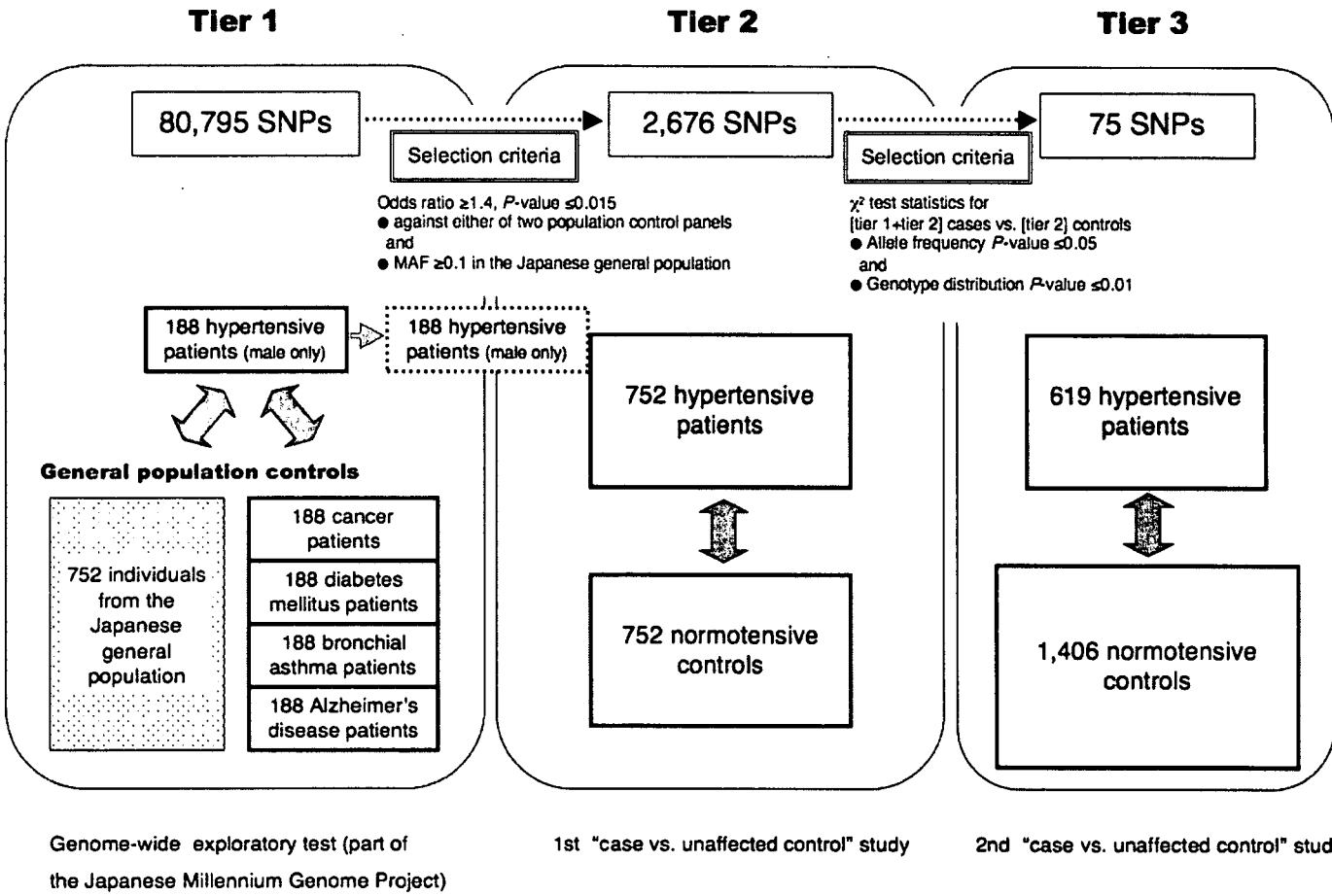

Figure 1. Schematic presentation of a three-tiered screening strategy in the present study. Gene-centered genome-wide exploratory test was performed in tier 1, followed by case–control study of disease-associated SNPs in tier 2 and tier 3 samples. A panel of 188 male hypertensive patients were compared with each of two population control panels in tier 1. Subsequently, 'case versus unaffected control' study was repeated twice to identify the best candidate SNPs. In transitions from tier 1 to tier 2, and from tier 2 to tier 3, the number of SNPs was reduced according to the selection criteria that we arbitrarily defined. See details in the Materials and Methods section.

the second 'case versus unaffected control' study (Fig. 2). Of these, we found six SNPs that showed *P*-values of ≤ 0.05 for both genotype distribution and allele frequency in the χ^2-test statistic (Table 3). rs3755351 and rs3771426 were located within the assumed intron 1 of *ADD2*, rs3787240 and rs3787241 were located in the same intron of *EYA2*, and the remaining two SNPs—rs3794260 and rs1805762—each located in *KIAA0789* and *M6PR*. To adjust for three covariates—age, gender and body mass index (BMI), we also performed logistic regression analysis for the significant SNPs (Supplementary Material, Table S1). With consideration of genetic model consistency, an SNP (rs3755351) showed the strongest association in the identical model (an additive model by logistic regression analysis) among three tiers. Further details of the association results are described in the Discussion.

SNP discovery and further test of association in three selected genes

Because a group of SNPs from three genes, *ADD2*, *KIAA0789* and *M6PR*, were particularly noted for their significant association with hypertension (Table 3), we searched for potentially functional SNPs by re-sequencing the 5′- and 3′-untranslated regions, all exons and exon–intron borders of the individual loci, on the basis of the gene structure deposited in the human genome database (http://www.ncbi.nlm.nih.gov/). We detected a total of 74 SNPs—25 SNPs in *ADD2*, 40 SNPs in *KIAA0789* and 9 SNPs in *M6PR*—and thereby selected 25 tag SNPs for genotyping 2025 subjects in tier 3 (see Supplementary Material, Table S2A, B and C). Apart from four SNPs which had been already included in the JSNP screening marker set, we found four additional SNPs, two in *ADD2* (rs2024453 and rs10084293) and one each in *KIAA0789* (rs9739493) and *M6PR* (rs1805725), to be significantly associated with hypertension (Table 3). Thus, in each gene, we identified at least two SNPs showing modest evidence of association with hypertension ($P \leq 0.05$ level in tier 3) but these SNPs did not necessarily belong to the same linkage disequilibrium (LD) block (Fig. 3 and LD group in Supplementary Material, Table S2A, B and C). The analysis of haplotypes inferable from tag SNPs did not show more significant disease association than the analysis of individual SNPs in any of three genes tested (data not shown).

Consideration of study power and multiple testing

We first estimated a type I error probability for the three-tiered screening to be 6.8×10^{-5}: 0.036 for tier 1, 0.0009 for tiers 1

Table 1. Summary of SNPs genotyped in tier-1 screening and genome coverage estimated by HapMap data

Chromosome	From JSNP screening markers		JSNP overlapped with HapMap (overlap)	From HapMap data (Release 21, JPT)			
	Total SNPs in JSNP	Proportion of SNPs unique to JSNP		SNPs in close LD ($r^2 \geq 0.8$) with overlap	(HapMap total−NA) SNPs	Coverage estimate: SNPs in LD ($r^2 \geq 0.8$)/(HapMap total−NA) SNPs	Total SNPs in HapMap
1	8378	0.370	5281	26 236	113 362	0.231	139 002
2	7336	0.293	5189	28 763	123 447	0.233	160 546
3	5128	0.358	3290	17 494	91 985	0.190	125 160
4	3172	0.366	2010	11 998	74 080	0.162	114 809
5	4973	0.311	3427	18 432	91 206	0.202	122 243
6	6220	0.272	4527	26 182	110 532	0.237	134 177
7	5813	0.358	3731	18 241	81 727	0.223	99 808
8	2388	0.246	1800	12 644	80 400	0.157	111 953
9	2818	0.218	2203	12 358	70 091	0.176	91 908
10	3159	0.322	2141	13 552	81 103	0.167	100 771
11	3636	0.248	2735	14 953	75 147	0.199	95 905
12	3816	0.223	2964	15 188	73 983	0.205	89 436
13	1291	0.290	917	6717	48 622	0.138	75 956
14	2913	0.219	2275	11 583	50 769	0.228	62 203
15	2311	0.194	1863	9903	46 599	0.213	54 210
16	2677	0.268	1959	8227	43 415	0.189	51 865
17	3246	0.258	2408	10 050	38 550	0.261	41 725
18	1243	0.207	986	6370	41 494	0.154	56 203
19	3392	0.308	2346	7626	25 524	0.299	26 949
20	2588	0.402	1548	8784	41 725	0.211	45 582
21	1761	0.275	1276	5777	23 465	0.246	26 892
22	2536	0.280	1825	7569	24 402	0.310	25 077
Total	80 795	0.298	56 701	298 647	1 451 628	0.206	1 852 380

The numbers of SNPs genotyped in tier-1 screening are demonstrated for each chromosome. Genome coverage was assessed with the HapMap data from JPT ($n = 45$); that is, the proportion of HapMap SNPs showing high r^2 (≥ 0.8) to one of the SNPs genotyped in this study (which are all derived from JSNP) is calculated. Because substantial part of the SNPs have turned out to be unique to JSNP, those overlapping with the HapMap SNPs, in the 'overlap' column, are used to estimate genome coverage. Here, NA represents a category of SNPs which have been mapped to the genome (NCBI B35) but do not have LD information against the HapMap SNPs. In this context, it is appropriate to reduce this NA SNPs from total SNPs deposited in the HapMap data when estimating genome coverage and we therefore use the number of SNPs (HapMap total−NA) as a denominator.

and 2 combined and 0.076 for tier 3 screening. Then, we estimated overall sensitivities (which could represent the statistical power) to be 0.10–0.45, 0.04–0.23 and 0.01–0.08 for a disease-associated SNP of OR = 1.4, 1.3 and 1.2, respectively, assuming the disease allele frequency within 0.1–0.9, the disease prevalence of 0.25 and the multiplicative genotype model. Since we had adopted relatively generous criteria for screening association signals, we evaluated the false discovery rate (FDR) to account for multiple testing (15). FDR for the nine SNPs found as significant was 0.69. A multi-staged screening in the current study could be largely categorized into two steps: tiers 1 and 2 (which constitute the first 'case versus unaffected control' study) and tier 3 (which constitutes the second 'case versus unaffected control' study). We therefore assessed experiment-wise type I errors with particular focus on the last-stage screening in tier 3. By permutation, the chance of observing a *P*-value of 0.0019 (for allele frequency test at rs3755351 in *ADD2*) in tier 3 was estimated to be 0.13.

DISCUSSION

With the recent advent of high-throughput genotyping technologies and high-resolution maps of SNP markers, it is expected that genome-wide association studies allow us to identify systematically the contributions of common genetic variations to human multifactorial diseases (16–18). In this line, our study has attempted to discover common hypertension susceptibility gene variants via a gene-centered genome-wide association design for the first time. Despite the modest genetic impacts assumed for hypertension, e.g. the λ-values (the relative risk for siblings of the affected probands) have been reported to be approximately 4 (19), we have nominated several susceptibility genes for hypertension (Table 3). Among these genes, findings for *ADD2* and *KIAA0789* are particularly noteworthy, because the former has been known to be a physiological candidate gene for hypertension and the latter is a novel gene with as-yet unknown physiological function.

Through a multi-tiered screening, nine SNPs derived from seven distinct gene loci have remained to show some evidence of association out of the 80 795 SNPs initially screened. Although the selection criteria were arbitrarily defined in the present study, a small percentage of the SNPs have passed the criteria in transitions from tier 1 to tier 2 (3.3%) and from tier 2 to tier 3 (2.8%). In the *ADD2* gene, for example, the minor allele frequency (MAF) of rs3755351 is lower in case groups (0.14–0.19) than that in control groups (0.21–0.22) throughout three tiers. A *P*-value of 1.7×10^{-5} and an OR of 1.30 (95% CI 1.15–1.46) are attained for allele frequency comparison of rs3755351 when the subjects studied in different tiers are combined and finally categorized into

Table 2. Clinical characteristics of participants

Variables	Case group Tier 2 panel	Tier 3 panel	Control group Tier 2 panel	Tier 3 panel
Number of subjects (female/male)	752 (353/399)	619 (280/339)	752 (366/386)	1406 (650/756)
Present age, year	62.4 ± 10.3	54.1 ± 8.4[a]	62.0 ± 8.7	58.4 ± 6.6
Age of onset, year	47.3 ± 10.2	43.2 ± 9.9	—	—
Current BMI, kg/m^2	23.9 ± 3.2	25.1 ± 3.6[a]	22.5 ± 2.8	22.4 ± 2.7
Smoking[b]				
None, %	48.6	61.6	66.0	58.6
Previous smoker, %	—	17.0	—	10.2
Current smoker, %	51.4	21.4	34.0	31.2
Blood pressure				
Systolic blood pressure, mmHg	146.4 ± 19.5[a]	150.9 ± 19.3[a]	113.8 ± 9.8	114.4 ± 10.1
Diastolic blood pressure, mmHg	86.4 ± 13.0[a]	91.4 ± 12.2[a]	69.8 ± 7.7	70.3 ± 7.2
Treatment of hypertension, %	92.6	75.4	—	—
Blood chemistry				
Serum creatinine, mg/dl	0.87 ± 0.69[a]	0.75 ± 0.50[c]	0.73 ± 0.18	0.70 ± 0.23
Fasting plasma glucose, mg/dl	105.3 ± 28.7	109.0 ± 31.0[a]	104.0 ± 41.9	99.2 ± 22.7
Serum total cholesterol, mg/dl	204.4 ± 31.1[c]	213.4 ± 33.8	209.2 ± 33.7	215.6 ± 34.1
Serum triglyceride, mg/dl	129.8 ± 82.2[a]	141.3 ± 124.6[a]	108.1 ± 67.1	110.1 ± 71.6
Serum HDL cholesterol, mg/dl	56.2 ± 16.7[a]	61.9 ± 19.5	60.6 ± 16.0	63.5 ± 17.5

Values are means ± SD.
For some variables, subjects with insufficient information are not included in the calculation.
[a]$P < 0.001$, case group versus control group by the unpaired t-test in each tier.
[b]Because of differences in the questionnaire, smoking status is categorized into two groups (non-smoker or smoker) in the tier 2 panel.
[c]$P < 0.01$, case group versus control group by the unpaired t-test in each tier.

the case (tiers 1–3) and unaffected control (tiers 2 and 3) groups. None of our results appears to be significant with the use of a strict Bonferroni correction, a very conservative evaluation of significance, and further replication in an independent population is indispensable.

The candidacy of *ADD2* as a hypertension susceptibility gene has been supported by several physiological and biochemical findings (20–22), together with some evidence from the studies of molecular genetics (23–27). Adducin is a ubiquitously expressed membrane-skeleton heteromeric protein composed of different subunits, α-, β- and γ-subunits. It is known to play a substantial role in the regulation of membrane ion transport. Point mutations of the α- and β-adducins account for up to 50% of the blood pressure difference between Milan hypertensive and normotensive rat strains, probably via the modulation of the Na^+–K^+ ATPase activity (one of major Na^+-channels) in the kidney (23,24). In this line, of note is the fact that β-adducin-deficient mice show significant increases in systolic and diastolic blood pressures and pulse pressure (21). The human homolog of β-adducin spans over 100 kb on chromosome 2p13 and comprises 17 exons. It has been reported that a common SNP (rs4984) identified at position 1797 in exon15 is associated with an increased risk of hypertension under certain pathological conditions in European populations (25–27), whereas this SNP itself is not polymorphic in Asian populations (http://www.ncbi.nlm.nih.gov/SNP/). Also, it has to be noted that one previous study (28) showed significant evidence for hypertension linkage in the 2p13 region (a peak of 2.84 LOD at 93 cM), where the *ADD2* locus is exactly located among several positional candidate genes. Despite our investigation in the *ADD2* locus, we could not find either a clear LD block-like structure or potentially functional SNPs in the vicinity of three disease-associated SNPs (rs2024453, rs3755351 and rs3771426), which are located in the putative promoter region and intron 1, apart from rs10084293 located within an LD block of *ADD2* (Fig. 3). We have assessed the independence of multiple associated SNPs in *ADD2* by logistic regression analysis and have found that the observed association in this gene could be explained principally by the most significant SNP (rs3755351) (see Supplementary Explanation). Once these associations are validated in an independent study panel, further extensive searches of functional SNPs in the *ADD2* locus are warranted.

Our high-density association study has also highlighted the *KIAA0789* gene located on chromosome 12q23.3. This gene encodes a hypothetical protein, LOC9671, which is expressed principally in the central nervous system and modestly in the pancreas (unpublished data). The predicted gene structure of *KIAA0789* involves 9 exons, spanning ~120 kb. There is a clear LD block in the 5′ region of the putative exon 1 (~3.8 kb in size), whereas we have found two other LD block-like structures within the *KIAA0789* gene (Fig. 3). Two disease-associated SNPs (rs3794260 and rs9739493) have turned out to reside in different LD blocks, and the construction of their haplotypes does not seem to provide much additional information on disease association. Although the precise gene structure and gene function remain unknown, *KIAA0789* appears to contain a carboxy-binding WSC domain, and its homologs are likely to exist in mice and rats according to the database information (http://www.ncbi.nlm.nih.gov/). Again, detailed investigation including independent

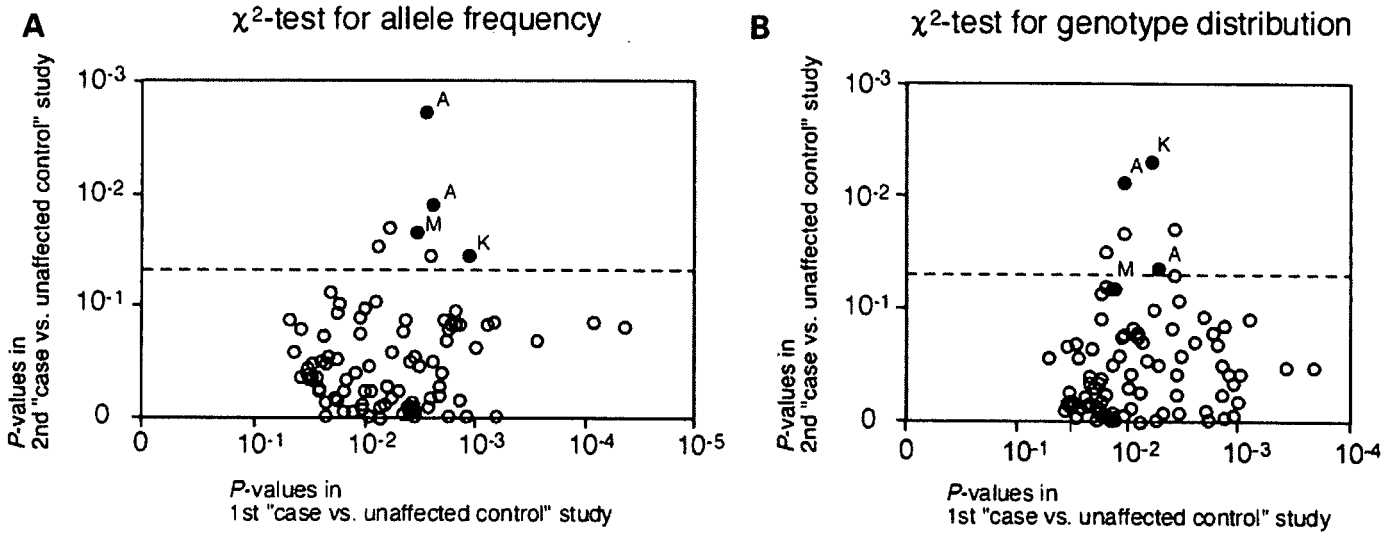

Figure 2. Statistical significance of χ^2-test of disease association in the two-staged 'case versus unaffected control' study. $-\mathrm{Log}_{10}$ *P*-values were used instead of raw *P*-values in each scatter plot. The dashed lines indicate $P = 0.05$. (**A**) As for SNPs genotyped in the second 'case versus unaffected control' panel, *P*-values for allele frequency in the second 'case versus unaffected control' panel were plotted against those in the first 'case versus unaffected control' panel, where SNPs located in three gene loci—*ADD2*, *KIAA0789* and *M6PR*—are depicted with solid circles to which the following symbols are attached: A, *ADD2*; K, *KIAA0789*; M, *M6PR*. (**B**) As for SNPs genotyped in the second 'case versus unaffected control' panel, *P*-values for genotype distribution ([2 × 3] contingency table) in the second 'case versus unaffected control' panel were plotted against those in the first 'case versus unaffected control' panel, where SNPs located in three gene loci—*ADD2*, *KIAA0789* and *M6PR*—are depicted with solid circles as mentioned earlier.

replication of disease association will lead us to clarify the etiological relevance of *KIAA0789* to hypertension.

Another, potential disease association, though modest statistical significance, has been found for *M6PR*. The *M6PR* gene encodes a cation-dependent receptor for mannose-6-phosphate groups on lysosomal enzymes and plays a critical role in the segregation and targeting of lysosomal enzymes to lysosomes. Thus far, no functional relation between *M6PR* and hypertension has been reported. Similar to *KIAA0789*, this gene could also allow us to identify a novel, as-yet unnoticed blood pressure regulatory mechanism.

We should bear in mind several limitations inherent in the present study. First, the level of genome coverage is an issue of heated debate (10,18). Some people may argue that our *a priori* marker selection strategy is gene-centric without utilizing LD information and hence it is not sufficient to pick up as many modest associations as possible in genome-wide searches of hypertension susceptibility genes. A comprehensive framework of common variations throughout the human genome has been made available by the recent completion of the International HapMap Project (29). On the basis of our assessment, the JSNP screening markers in this study cover 20.6% of the HapMap SNPs, whereas a substantial proportion (~30%) of SNPs appear to be unique to JSNP (Table 1 and Supplementary Material, Fig. S1). Under these circumstances, an ideal set of SNPs for our study would encompass deliberately selected tag SNPs (principally common genetic variants) and additional 'singleton' SNPs (sometimes rare genetic variants). Besides this argument of tag SNPs, there are two points of weakness regarding genome coverage as follows: (i) sex chromosome markers have been excluded from the analysis because of the pre-determined policy of multi-disease collaborative study in the Japanese Millennium Genome Project, and (ii) a substantial part of the expressed human genes is not covered by the JSNP database (11), in which the fundamental SNP data were almost fixed in the middle of 2003. Second, the statistical power attainable by our study panel needs to be taken into consideration. For the last few years, genotype costs have fallen dramatically, yet present economic and experimental conditions make it necessary, in practice, to reduce the number of genotyped samples down to a moderately sized case group (188 subjects in our study) at the initial screening with approximately 80 000 SNPs. We arbitrarily set the selection criteria of $\mathrm{OR} \geq 1.4$ and $P \leq 0.015$ in transition from tier 1 to tier 2, where the overall statistical power is estimated to be 10–45% for a disease-associated SNP of $\mathrm{OR} = 1.4$ and 1–8% for that of $\mathrm{OR} = 1.2$, assuming the disease allele frequency within 0.1–0.9 and the disease prevalence of 0.25. Thus, it is likely that our study design allows for capturing less than half of the true disease associations particularly with regard to modest genetic susceptibility. Third, ethnic diversity has not been tested within the scope of the present study. Instead of using commercially available SNP sets aimed at full genomic coverage, we have attempted to focus on potentially functional variants and also relatively common SNPs ($\mathrm{MAF} \geq 0.1$) in the Japanese population. Accordingly, some of disease-associated SNPs listed in Table 3 may be rare or not polymorphic in the other ethnic groups. To clarify allele frequency representation of individual loci and etiological impacts attributable to them, further examination is required in the context of ethnic diversity.

During our preparation of this report, two genome-wide association studies for hypertension and/or blood pressure have been performed in Caucasians (30,31). When our results are compared with public data sets for these association statistics, a few SNPs in the regions of interest appear to show a tendency of association with hypertension or blood pressure;

Table 3. Summary of genomic SNPs associated with hypertension status in two-staged 'case versus unaffected control' study

dbSNP number	Gene symbol	Major/ minor allele	Orientation	Minor allele frequency						Association analysis												
				Case			Control			*P*-value in the first 'case (tiers 1 and 2) versus unaffected control (tier 2)' study				*P*-value in the second 'case (tier 3) versus unaffected control (tier 3)' study				Total subjects: case (tiers 1–3) versus unaffected control (tiers 2 and 3)				
				Tier 1	Tier 2	Tier 3	Tier 1[a] (Ref. 1/ Ref. 2)	Tier 2	Tier 3	[2 × 3] Contingency	Dominant	Recessive	Allele	[2 × 3] Contingency	Dominant	Recessive	Allele	[2 × 3] Contingency *P*-value	Dominant *P*-value	Recessive *P*-value	Allele frequency model *P*-value	Allele frequency model OR (95% CI)[b]
rs3755351	ADD2	C/A	fwd	0.14	0.19	0.17	(0.22/0.21)	0.22	0.21	0.011	0.041	0.007	0.003	0.007	0.071	0.003	0.002	0.00009	0.009	0.00006	0.00002	1.30 (1.15–1.46)
rs3771426	ADD2	T/C	fwd	0.13	0.16	0.16	(0.19/NA)	0.2	0.19	0.005	0.006	0.016	0.003	0.043	0.093	0.025	0.012	0.0003	0.003	0.0007	0.00008	1.28 (1.13–1.45)
rs2024453	ADD2	T/C	rev	—	—	0.29	—	—	0.33	—	—	—	—	0.034	0.013	0.132	0.021	—	—	—	—	—
rs10084293	ADD2	G/A	rev	—	—	0.44	—	—	0.48	—	—	—	—	0.089	0.083	0.063	0.027	—	—	—	—	—
rs3794260	KIAA0789	G/A	fwd	0.14	0.17	0.17	(0.21/0.18)	0.21	0.20	0.006	0.161	0.002	0.001	0.005	0.001	0.256	0.035	0.0008	0.006	0.001	0.0001	1.26 (1.12–1.42)
rs9739493	KIAA0789	T/C	fwd			0.41			0.44					0.026	0.008	0.641	0.041					
rs1805762	M6PR	C/G	rev	0.21	0.22	0.22	(0.25/0.24)	0.26	0.25	0.015	0.045	0.009	0.003	0.061	0.332	0.019	0.022	0.001	0.615	0.0003	0.0003	1.23 (1.10–1.37)
rs1805725	M6PR	T/G	fwd			0.54	–		0.49					0.007	0.002	0.123	0.005					
rs3787240	EYA2	C/T	fwd	0.25	0.21	0.22	(0.19/0.18)	0.18	0.19	0.003	0.001	0.059	0.008	0.083	0.227	0.034	0.028	0.001	0.002	0.007	0.0010	0.82 (0.73–0.92)
rs3787241	EYA2	G/A	fwd	0.24	0.21	0.22	(0.19/0.19)	0.18	0.19	0.008	0.003	0.080	0.014	0.104	0.267	0.042	0.037	0.002	0.003	0.01	0.0017	0.83 (0.74–0.93)
rs3761987	—	T/A	fwd	0.43	0.37	0.37	(0.36/0.36)	0.33	0.34	0.004	0.001	0.101	0.004	0.019	0.566	0.015	0.134	0.0007	0.031	0.004	0.0016	0.86 (0.78–0.94)
rs3741691	THAP2	A/C	fwd	0.26	0.22	0.22	(0.21/0.20)	0.19	0.21	0.011	0.043	0.007	0.002	0.021	0.139	0.066	0.317	0.007	0.727	0.002	0.0064	0.85 (0.76–0.96)
rs1298463	CCDC131	A/G	fwd	0.26	0.22	0.22	(0.21/0.19)	0.19	0.21	0.016	0.052	0.010	0.004	0.031	0.182	0.073	0.312	0.012	0.736	0.0033	0.0096	0.86 (0.77–0.96)

Two SNPs of EYA2, rs3787240 and rs3787241, are located closely (only 296 bp apart) and in complete LD ($r^2 = 1.00$) to each other.

Also, rs3741691 and rs1298463 are located closely (44 kb apart) and have turned out to be in strong LD ($r^2 = 0.99–1.00$) to each other.

[a]In tier 1 control subjects, the figures in parentheses are minor allele frequencies (MAFs) calculated separately in the Ref. 1 panel—the other disease patients who can be regarded as arbitrary general controls; Ref. 2 panel—752 individuals from the Japanese general population (see Materials and Methods).

[b]The OR was calculated as the ratio of the odds of disease in chromosomes with major alleles relative to those without them.

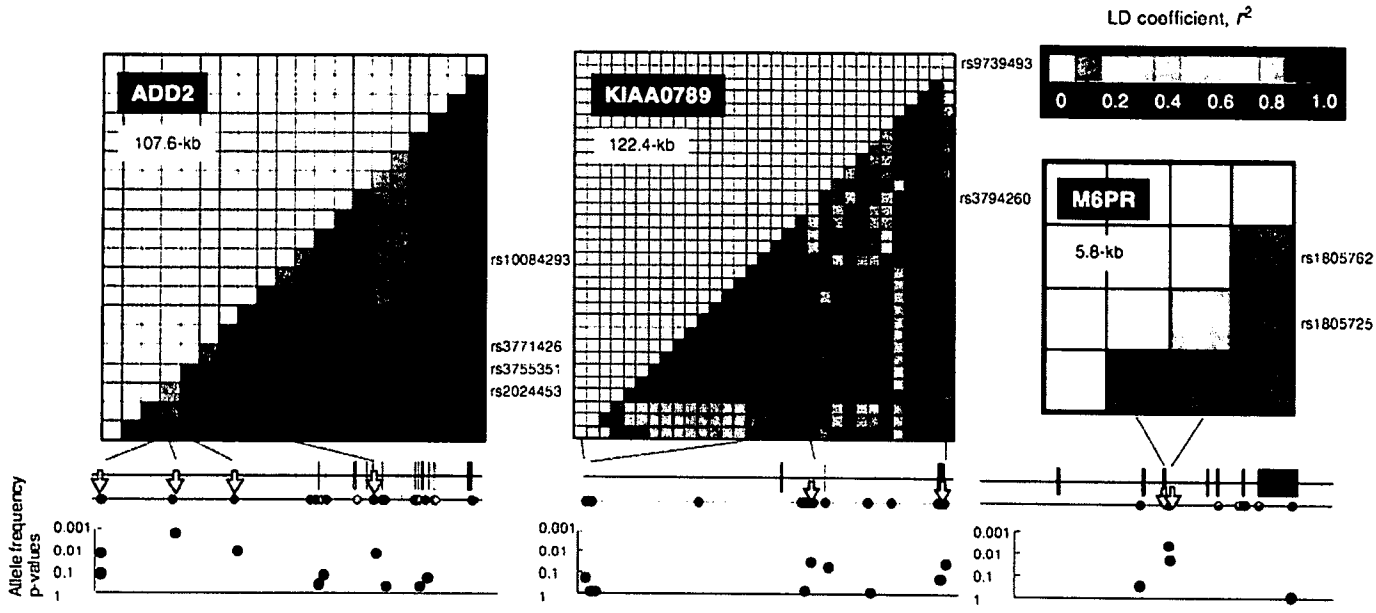

Figure 3. LD relations between SNPs in the *ADD2*, *KIAA0789* and *M6PR* genes (top) and disease association of markers from the corresponding genomic regions (bottom). In the top, the LD between a pair of markers is indicated by the color of the block above and to the left of the intersection of the markers. For the sake of readability, only the names of SNPs showing significant association are shown to the right of the vertical axis of the LD plot. The rest of the SNP information is described in Supplementary Material, Table S2A, B and C. In the upper bottom, the location of genetic markers studied in the corresponding genomic region is shown with relation to gene structure. Here, green and red circles indicate the SNPs with low (MAF < 0.05) and high (MAF ≥ 0.05) allele frequencies in the Japanese, respectively. In the lower bottom, $-\log_{10}$ *P*-values of the differences in allele frequencies between case and control subjects in tier 3 (i.e. the second 'case versus unaffected control' study) were plotted against the location of individual genetic markers genotyped.

for example, an SNP (rs17006246) in *ADD2*, which is in strong LD ($r^2 = 0.806$ and $D' = 1$ in the HapMap JPT population) with rs3755351, the most significant SNP in our study, is modestly associated with hypertension status ($P = 0.029$) in the Diabetes Genetics Initiative (DGI) study but the direction of effect is opposite between rs3755351 typed in this study and rs17006246 typed in the DGI study. On the other hand, rs1805740, in strong LD with an SNP (rs1805762) in *M6PR*, is modestly associated with hypertension status ($P = 0.036$) in the Wellcome Trust Case Control Consortium study with the same direction of effect as in this study (see Supplementary Material, Table S3).

In summary, our high-density association study provides a list of gene loci potentially predisposing people to hypertension, which awaits replication across populations. With the available samples, we have observed an association of SNPs including three SNPs clusters (or gene loci) in the Japanese populations. In face of the complex nature of disease etiology, it seems to be a formidable task but worth challenging that we eventually apply the SNPs information to improved prevention, diagnosis and treatment of hypertension.

MATERIALS AND METHODS

Study design

We performed a large-scale association study for genes susceptible to hypertension by using a three-tiered genotyping approach (tiers 1, 2 and 3) as depicted in Figure 1. All methods of the study were approved by the review committees of the individual institutions involved in the present study. All subjects provided written informed consent for participation.

In the gene-centered genome-wide exploratory test in tier 1, we carried out genotyping of 83 802 SNPs (3007 of which were excluded from the analysis because they are on sex chromosomes or in the unknown locations) using genomic DNAs from 188 Japanese male hypertensive patients and 752 unrelated Japanese individuals (referred to as general population controls) and another panel of 752 Japanese subjects (referred to as arbitrarily defined controls) who were affected with any of the other four common diseases including gastric cancer, diabetes mellitus, bronchial asthma and Alzheimer's disease; each of these was investigated as the 'Japanese Millennium Genome Project' (Fig. 1). The theoretical basis of adopting this exploratory test scheme was previously reported elsewhere (32). Cases were enrolled from the clinical practice or the annual medical checkup of university hospitals and medical institutions according to the uniformly defined criteria. These included (i) systolic blood pressure ≥ 160 mmHg, diastolic blood pressure ≥ 95 mmHg, or both on two consecutive visits for untreated subjects; (ii) patients receiving long-term antihypertensive treatments; (iii) no secondary form of hypertension as evaluated by an extensive workup; (iv) family history of hypertension, i.e. at least one hypertensive subjects detectable among parents and siblings of the participants; (v) an age of onset known to be between 30 and 59 years. Moreover, only male subjects with BMI < 25 kg/m^2 were selected in tier 1. We compared allele frequencies and/or genotype distributions in hypertensive patients and two population control panels and evaluated deviation from HWE at each of the genotyped loci. For the subsequent screening

in tier 2, we selected SNPs (i) with OR $\geq$ 1.4 and $P \leq 0.015$ against either of two population control panels and with concordant OR tendency against two control panels; (ii) with MAF $\geq$ 0.1 and (iii) not showing significant deviations ($P = 0.01$ level) from Hardy–Weinberg expectations in the patient or control panels.

In tier 2 (which comprised 752 hypertensive patients and 752 normotensive controls), we further tested the SNPs thus screened in tier 1, which effectively constituted the first 'case (tiers 1 and 2) versus unaffected control (tier 2)' study. Here, cases in tier 2 were selected according to the criteria (i)–(v) mentioned earlier for tier 1. Normotensive controls, on the other hand, were defined as follows: (i) systolic blood pressure $\leq$130 mmHg and diastolic blood pressure $\leq$85 mmHg without receiving antihypertensive treatments; (ii) age $\geq$50 years and (iii) no family history of hypertension. Both males and females were included in tier 2 without reference to BMI. We selected SNPs (i) with P-value $\leq$0.05 when comparing allele frequency; and (ii) with P-value $\leq$0.01 when comparing genotype distribution between (tiers 1 and 2) cases and (tier 2) controls by χ^2 test statistics.

In tier 3 (which comprised 619 hypertensive patients and 1406 normotensive controls), we performed the second 'case versus unaffected control' study to examine significant associations observed in tiers 1 and 2. The diagnostic criteria in tier 3 were identical to those in tier 2. For the assessment of assumptions when using statistical models in the present study, quantile–quantile plots of P-values were depicted for each stage of association test described in Supplementary Explanation.

No significant population stratification was observed for samples in tier 1 when it was assessed with the methods reported by Patterson *et al.* (33). However, the presence of population stratification was indicated for samples in the first stage 'case (tiers 1 and 2) versus unaffected control' study. We observed moderate bias in genotype frequency of some SNPs between the two tiers, which may have resulted from technical/experimental artifacts between genotyping of cases in tiers 1 and 2. Therefore, the trend test statistic at this analytical stage was corrected according to the significant eigenvector (see Supplementary Explanation). Stratification in tier 3 was not detected but could not be ruled out because of the relatively small number of SNPs ($n = 75$) genotyped in tier 3. As for the nine SNPs that showed significant disease association after multi-stage screening, they were not correlated with the significant eigenvector detected in tiers 1 and 2 cases and tier 2 controls. The P-values for nine SNPs were similar between the nominal and the EIGENSTRAT-corrected ones; for example, the nominal P-value was 0.0029 and the EIGENSTRAT-corrected P-value was 0.0069 at rs3755351 in *ADD2*.

SNP marker resource and genotyping

Most of the SNP markers used in the present study were same as the markers used in the previous reports (14) and derived from the JSNP database. The samples in tiers 1 and 2 were genotyped by PCR amplification of multiple genomic fragments with 20 ng of genomic DNA followed by characterization with the invader assay. Genotyping of the samples in tier 3 was undertaken using the TaqMan® SNP Genotyping Assays (Applied Biosystems). To secure the accuracy and completeness of genotyping, which is critical for large-scale studies (34), we attached a set of 'flags' to individual SNP data mainly dependent on the data completeness, after two independent investigators had checked the raw data robustness by looking at the scatter plot of the assay.

SNP discovery in the selected genes

Approximately 38 kb of genomic sequence spanning the exons and the 5′- and 3′-untranslated regions of three genes, *ADD2*, *KIAA0789* and *M6PR*, was re-sequenced in 48 Japanese control individuals to identify potentially functional SNPs. Since *KIAA0789* had not been fully annotated, the arbitrary positions of translation initiation sites were estimated according to the human genome database. From the SNPs thus identified, tag SNPs were selected for the three genes with the algorithm that we previously reported (35). These tag SNPs were then used for the case–control analysis in tier 3 to further examine association signals seen throughout the multi-staged screening. We deposited the identified SNP information in the NCBI's SNP database and also in our own database, JMDBase (Japan Metabolic Disease Database).

Statistical analysis

The SNPs were tested individually for the statistical significance of disease association with the χ^2-test statistic, which evaluated three inheritance models—[2 × 3] contingency table, dominant and recessive models—for genotype distributions and independence on [2 × 2] contingency table for allele frequencies. Here, the most significant P-values among three inheritance models were adopted for genotype distributions when we selected SNPs for screening in tier 3. The criteria for declaring suggestive evidence of disease association were arbitrarily set at each analytical stage as summarized in Figure 1, and they are described in the Results section. SNPs' genotype departures from HWE were tested using the χ^2-test with 1 degree of freedom.

In the three genes showing significant association signals, the extent of LD was measured in terms of an LD coefficient r^2 before the analysis of haplotype structure. Within each LD block, haplotypes were inferred from genotype data by the SNPHAP software for the case and control groups, respectively.

We randomly permutated the genotype of individuals across different panels, 100 times per SNP, and counted the ratio of permutations that fulfill the screening criteria. This ratio indicates the specificity of the study. According to the P-value distribution of the permutations, we evaluated the probability of observing an SNP with P-value no larger than the actual minimum. This probability indicates the experiment-wise P-value. For the specific prevalence and penetrance, we calculated genotype frequency and randomly generated genotypes according to their frequency. We generated genotypes for 1000 simulations of each panel and computed the ratio of simulations that could pass the screening. This ratio is considered the sensitivity of the study.

Values were expressed as means ± SD unless otherwise indicated.

Uniform resource locators

The JSNP database is available at http://snp.ims.u-tokyo.ac.jp/index.html. The National Center for Biotechnology Information's SNP database is available at http://www.ncbi.nlm.nih.gov/SNP/. The JMDBase is available at http://www.jmdbase.jp. SNPHAP is available at http://www.gene.cimr.cam.ac.uk/clayton/software/

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We acknowledge the outstanding contributions of the SNP Research Center, the Institute of Physical and Chemical Research (RIKEN), the National Cancer Center (NCC) and the International Medical Center of Japan (IMCJ) employees who provided technical and infrastructural support for this work. Above all, we thank the patients and study subjects who made this work possible and who give it value. We thank all the people who continuously support the Suita Study. We also thank Drs Yukio Yamori, Yoshihiro Kokubo, Masakuni Serizawa, Ken Sugimoto, Hiroshi Akasaka, Chikanori Makibayashi and many physicians in Amagasaki Medical Association, Sadahiko Iwamoto, Yoichi Ohno, Mikano Hishiki, Kotoko Kosuge, Noriko Aoi, Akira Haketa, Yoichi Izumi, Toru Izumi, Taishi Sasaoka, Shigeru Hohda and Takeharu Hayashi, and Ms Kaoru Sugama for their contribution to collecting DNA samples and clinical information accompanied; Drs Hiromi Rakugi and Masafumi Kuremura and Ms Akemi Muramoto and Kazuko Iwasa for their advice and assistance.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Grant-in-Aid for Scientific Research on Priority Areas (Millennium Genome Project) from the Ministry of Education, Science, Sports, and Culture of Japan, the grants from the Program for Promotion of Fundamental Studies in Health Sciences of Pharmaceuticals and Medical Devices Agency (PMDA) and a grant for the Core Research for the Evolutional Science and Technology, from the Japan Science Technology Agency (JST).

REFERENCES

1. Mitchell, G.F., DeStefano, A.L., Larson, M.G., Benjamin, E.J., Chen, M.H., Vasan, R.S., Vita, J.A. and Levy, D. (2005) Heritability and a genome-wide linkage scan for arterial stiffness, wave reflection, and mean arterial pressure: the Framingham Heart Study. *Circulation*, **112**, 194–199.
2. Williams, R.R., Hunt, S.C., Hasstedt, S.J., Hopkins, P.N., Wu, L.L., Berry, T.D., Stults, B.M., Barlow, G.K., Schumacher, M.C., Lifton, R.P. *et al.* (1991) Are there interactions and relations between genetic and environmental factors predisposing to high blood pressure? *Hypertension*, **18** (Suppl 3), I29–I37.
3. Biron, P., Mongeau, J.G. and Bertrand, D. (1976) Familial aggregation of blood pressure in 558 adopted children. *Can. Med. Assoc. J.*, **115**, 773–774.
4. Gong, M. and Hubner, N. (2006) Molecular genetics of human hypertension. *Clin. Sci.*, **110**, 315–326.
5. Rice, T., Cooper, R.S., Wu, X., Bouchard, C., Rankinen, T., Rao, D.C., Jaquish, C.E., Fabsitz, R.R. and Province, M.A. (2006) Meta-analysis of genome-wide scans for blood pressure in African American and Nigerian samples. The National Heart, Lung, and Blood Institute GeneLink Project. *Am. J. Hypertens.*, **19**, 270–274.
6. Liu, W., Zhao, W. and Chase, G.A. (2004) Genome scan meta-analysis for hypertension. *Am. J. Hypertens.*, **17**, 1100–1106.
7. Province, M.A., Kardia, S.L., Ranade, K., Rao, D.C., Thiel, B.A., Cooper, R.S., Risch, N., Turner, S.T., Cox, D.R., Hunt, S.C. *et al.* (2003) National Heart, Lung and Blood Institute Family Blood Pressure Program. *Am. J. Hypertens.*, **16**, 144–147.
8. Jones, H.B. (1998) The relative power of linkage and association studies for the detection of genes involved in hypertension. *Kidney. Int.*, **53**, 1446–1448.
9. Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science*, **273**, 1516–1517.
10. Barrett, J.C. and Cardon, L.R. (2006) Evaluating coverage of genome-wide association studies. *Nat. Genet.*, **38**, 659–662.
11. Hirakawa, M., Tanaka, T., Hashimoto, Y., Kuroda, M., Takagi, T. and Nakamura, Y. (2002) JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res.*, **30**, 158–162.
12. Haga, H., Yamada, R., Ohnishi, Y., Nakamura, Y. and Tanaka, T. (2002) Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190,562 genetic variations in the human genome. Single-nucleotide polymorphism. *J. Hum. Genet.*, **47**, 605–610.
13. Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. and Nakamura, Y. (2001) A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.*, **46**, 471–477.
14. Yamazaki, K., McGovern, D., Ragoussis, J., Paolucci, M., Butler, H., Jewell, D., Cardon, L., Takazoe, M., Tanaka, T., Ichimori, T. *et al.* (2005) Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum. Mol. Genet.*, **14**, 3499–3506.
15. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.*, **57**, 289–300.
16. Wang, W.Y., Barratt, B.J., Clayton, D.G. and Todd, J.A. (2005) Genome-wide association studies: theoretical and practical concerns. *Nat. Rev. Genet.*, **6**, 109–118.
17. Hirschhorn, J.N. and Daly, M.J. (2005) Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.*, **6**, 95–108.
18. Palmer, L.J. and Cardon, L.R. (2005) Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet*, **366**, 1223–1234.
19. Altmuller, J., Palmer, L.J., Fischer, G., Scherb, H. and Wjst, M. (2001) Genomewide scans of complex human diseases: true linkage is hard to find. *Am. J. Hum. Genet.*, **69**, 936–950.
20. Gilligan, D.M., Lozovatsky, L. and Silberfein, A. (1997) Organization of the human beta-adducin gene (ADD2). *Genomics*, **43**, 141–148.
21. Marro, M.L., Scremin, O.U., Jordan, M.C., Huynh, L., Porro, F., Roos, K.P., Gajovic, S., Baralle, F.E. and Muro, A.F. (2000) Hypertension in beta-adducin-deficient mice. *Hypertension*, **36**, 449–453.
22. Muro, A.F., Marro, M.L., Gajovic, S., Porro, F., Luzzatto, L. and Baralle, F.E. (2000) Mild spherocytic hereditary elliptocytosis and altered levels of alpha- and gamma-adducins in beta-adducin-deficient mice. *Blood*, **95**, 3978–3985.
23. Tripodi, G., Valtorta, F., Torielli, L., Chieregatti, E., Salardi, S., Trusolino, L., Menegon, A., Ferrari, P., Marchisio, P.C. and Bianchi, G. (1996) Hypertension-associated point mutations in the adducin alpha and beta subunits affect actin cytoskeleton and ion transport. *J. Clin. Invest.*, **97**, 2815–2822.
24. Ferrandi, M., Salardi, S., Tripodi, G., Barassi, P., Rivera, R., Manunta, P., Goldshleger, R., Ferrari, P., Bianchi, G. and Karlish, S.J. (1999) Evidence for an interaction between adducin and Na(+)–K(+)-ATPase: relation to genetic hypertension. *Am. J. Physiol.*, **277**, H1338–H1349.
25. Tikhonoff, V., Kuznetsova, T., Stolarz, K., Bianchi, G., Casiglia, E., Kawecka-Jaszcz, K., Nikitin, Y., Tizzone, L., Wang, J.G. and

Staessen, J.A. (2003) beta-Adducin polymorphisms, blood pressure, and sodium excretion in three European populations. *Am. J. Hypertens.*, **16**, 840–846.
26. Tikhonoff, V., Kuznetsova, T., Stolarz, K., Bianchi, G., Casiglia, E., Kawecka-Jaszcz, K., Nikitin, Y., Tizzone, L., Wang, J.G., Staessen, J.A. and European Project on Genes in Hypertension Investigators (2003) Blood pressure phenotypes in relation to the beta-adducin C1797T polymorphism in the European Project on Genes in Hypertension (EPOGH). *Blood. Press. Monit.*, **8**, 151–154.
27. Wang, J.G., Staessen, J.A., Barlassina, C., Fagard, R., Kuznetsova, T., Struijker-Boudier, H.A., Zagato, L., Citterio, L., Messaggio, E. and Bianchi, G. (2002) Association between hypertension and variation in the alpha- and beta-adducin genes in a white population. *Kidney Int.*, **62**, 2152–2159.
28. Barkley, R.A., Chakravarti, A., Cooper, R.S., Ellison, R.C., Hunt, S.C., Province, M.A., Turner, S.T., Weder, A.B., Boerwinkle, E. and Family Blood Pressure Program (2004) Positional identification of hypertension susceptibility genes on chromosome 2. *Hypertension*, **43**, 477–482.
29. International HapMap Consortium (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
30. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
31. Diabetes Genetics Initiative of Broad Institute of Harvard MIT, Lund University, and Novartis Institutes of BioMedical ResearchSaxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N. *et al.* (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, **316**, 1331–1336.
32. Sato, Y., Suganami, H., Hamada, C., Yoshimura, I., Yoshida, T. and Yoshimura, K. (2004) Designing a multistage, SNP-based, genome screen for common diseases. *J. Hum. Genet.*, **49**, 669–676.
33. Patterson, N., Price, A.L. and Reich, D. (2006) Population structure and eigenanalysis. *PLoS Genet.*, **2**, e190.
34. Clayton, D.G., Walker, N.M., Smyth, D.J., Pask, R., Cooper, J.D., Maier, L.M., Smink, L.J., Lam, A.C., Ovington, N.R., Stevens, H.E. *et al.* (2005) Population structure, differential bias and genomic control in a large-scale, case–control association study. *Nat. Genet.*, **37**, 1243–1246.
35. Takeuchi, F., Yanai, K., Morii, T., Ishinaga, Y., Taniguchi-Yanai, K., Nagano, S. and Kato, N. (2005) Linkage disequilibrium grouping of single nucleotide polymorphisms (SNPs) reflecting haplotype phylogeny for efficient selection of tag SNPs. *Genetics*, **170**, 291–304.

ELSEVIER

Atherosclerosis 196 (2008) 29–36

ATHEROSCLEROSIS

www.elsevier.com/locate/atherosclerosis

Genetic variants in *PCSK9* in the Japanese population: Rare genetic variants in *PCSK9* might collectively contribute to plasma LDL cholesterol levels in the general population

Yasuko Miyake[a,*], Rina Kimura[a], Yoshihiro Kokubo[b], Akira Okayama[b], Hitonobu Tomoike[c], Taku Yamamura[d], Toshiyuki Miyata[a]

[a] *Department of Etiology and Pathophysiology, National Cardiovascular Center Research Institute, 5-7-1 Fujishirodai, Suita, Osaka 565-8565, Japan*
[b] *Division of Preventive Cardiology, National Cardiovascular Center, 5-7-1 Fujishirodai, Suita, Osaka 565-8565, Japan*
[c] *National Cardiovascular Center, 5-7-1 Fujishirodai, Suita, Osaka 565-8565, Japan*
[d] *Division of Health Sciences, Osaka University Graduate School of Medicine, 1-7 Yamadaoka, Suita, Osaka 565-0871, Japan*

Received 19 May 2006; received in revised form 21 November 2006; accepted 2 December 2006
Available online 21 February 2007

Abstract

The aim of this study was to investigate whether plasma low-density lipoprotein cholesterol (LDL-C) levels in the general population are influenced by rare sequence variations in the *PCSK9* gene. We sequenced the promoter and coding regions of the *PCSK9* gene in individuals from the general population ($n = 3655$) with the lowest ($n = 78$) and highest ($n = 96$) LDL-C levels and in individuals taking antihypercholesterolemia medication ($n = 96$). We identified 33 sequence variants in the *PCSK9* gene among which 24 were specific for Japanese. Statistical analysis showed that one missense mutation, R93C, was associated with low LDL-C levels. The other variants had no association with LDL-C levels or the numbers of individuals with the variants were too small for statistical analysis. A comparison of the numbers of individuals with nonsynonymous mutations between the low LDL-C and high LDL-C/treatment groups found that four missense mutations and one nonsense mutation were identified only in the low LDL-C group and six missense mutations were identified only in the high LDL-C/treatment group. As we have analyzed groups at opposite ends of the LDL-C spectrum, it is likely that some of these nonsynonymous mutations may be associated with either low or high LDL-C in the Japanese population. Based on the extremely high frequencies of the nonsynonymous mutations in *PCSK9* compared with those of *LDLR* or *apoB-100*, *PCSK9* mutations could be important factors that cumulatively influence plasma LDL-C levels in the general population.

Keywords: *PCSK9*; Plasma LDL cholesterol; Rare genetic variants; General population; Missense mutation

1. Introduction

Elevated plasma concentration of low-density lipoprotein cholesterol (LDL-C) is a major risk factor for the development and progression of atherosclerosis. Plasma concentrations of LDL-C are determined primarily by the activity of the LDL receptor (LDLR) in the liver. Recently, the *proprotein convertase subtilisin/kexin type 9* (*PCSK9*) gene has been found to be involved in the post-transcriptional regulation of the LDLR. *PCSK9* encodes a protein of 692 amino acids which is a member of the subtilisin-like protein convertase family [1,2] and is expressed most abundantly in the liver, kidneys and small intestine [2]. PCSK9 consists of several domains: a signal peptide, a prosegment, a subtilase-like catalytic domain and a C-terminal domain [3]. It is synthesized as a soluble zymogen which undergoes autocatalytic intramolecular cleavage in the endoplasmic reticulum (ER) between the prosegment and the catalytic domain [1,2].

* Corresponding author. Tel.: +81 6 6833 5012x2477; fax: +81 6 6872 8091.
E-mail address: ymiyake@ri.ncvc.go.jp (Y. Miyake).

0021-9150/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.atherosclerosis.2006.12.035

After cleavage, the mature PCSK9 exits the ER and is efficiently secreted [2]. The only known substrate of PCSK9 is itself; no other substrate(s) for PCSK9 have yet been identified. Although even the physiological substrate remains unknown, PCSK9 has been shown to play a role in cholesterol metabolism by regulating the number of cell-surface LDLRs [3–5].

Overexpression of the wild-type *Pcsk9* gene in mice results in hypercholesterolemia because of a reduced number of LDLRs [3–5]. The reduced number of LDLRs due to PCSK9 is not accompanied by changes in *LDLR* mRNA levels; therefore, it is likely that PCSK9 is involved in the post-transcriptional regulation of the LDLR [4,5]. Degradation of the LDLR by PCSK9 is dependent on the catalytic activity of PCSK9 [5,6]. In contrast, mice expressing no PCSK9 have markedly increased hepatic LDLR levels, resulting in accelerated LDL clearance [7]. These findings indicate that PCSK9 normally acts to limit the number of LDLRs at the cell surface. Thus, *PCSK9* mutations which disrupt normal function, i.e., loss-of-function mutations, are presumed to increase the number of LDLRs, resulting in hypocholesterolemia. In fact, the nonsense mutations identified by Cohen et al. are associated with a 40% reduction in mean plasma levels of LDL-C [8]. On the other hand, some mutations in the *PCSK9* gene cause hypercholesterolemia [9–11], which are probably due to gain-of-function mechanisms. These mutations in *PCSK9* might promote the degradation of LDLRs in hepatocytes [3–5]. Recently, Cameron et al. demonstrated that loss-of-function mutations in *PCSK9* increase the number of cell-surface LDLRs, while gain-of-function mutations decrease the number of LDLRs, based on studies on HepG2 cells transfected with mutant *PCSK9* constructs [12].

Since mutations in *PCSK9* can cause severe hypercholesterolemia [9–11] as well as hypocholesterolemia [8,13], sequence variants of *PCSK9* might contribute to variations in the plasma levels of LDL-C. Shioji et al. [14] have identified the two common single nucleotide polymorphisms (SNPs), and Chen et al. [15] identified a haplotype associated with differences in plasma LDL-C levels. Kotowski et al. performed a systematic examination of the relationship between sequence variations in *PCSK9* and plasma levels of LDL-C in the general population [16]. They analyzed sequence variations in *PCSK9* in individuals of the examined population who had lower and higher LDL-C levels and found that three missense mutations and two noncoding sequence variants were significantly associated with lower levels of LDL-C, while a single noncoding variant was associated with a modest increase in LDL-C levels. They concluded that sequence variants in *PCSK9* contribute significantly to interindividual variations in plasma LDL-C levels, and report that the spectrum of *PCSK9* alleles associated with LDL-C levels spanned a wide range of allele frequencies and magnitude of phenotypic effects.

In order to verify whether sequence variants in *PCSK9* could be a determinant of LDL-C plasma levels in the Japanese general population, we performed sequence analyses in the proximal promoter and all exons of *PCSK9* in individuals from the general population with the lowest and highest LDL-C levels and also in individuals taking antihypercholesterolemia medication since these individuals are presumed to have originally high levels of plasma LDL-C. Finally, we performed statistical analyses and compared the numbers of individuals with certain genetic variants between groups.

2. Methods

2.1. General population and the three investigated groups of individuals

DNA analysis was performed in individuals selected from the participants of the Suita cohort study, whose total sample included 3655 subjects. The study design of the Suita study has been described previously [17–19]. Briefly, the individuals were randomly selected from the municipal population registry, taking into consideration group stratification by gender and 10-year age divisions. The subjects visited the National Cardiovascular Center every 2 years for general health checkups. In addition to performing routine blood examinations, including lipid profiles, glucose levels, blood pressure and anthropometric measurements, a physician or nurse administered questionnaires regarding the individual's personal history of cardiovascular disease, including angina pectoris, myocardial infarction and stroke. Leukocyte DNA was collected from individuals who visited the National Cardiovascular Center between April 2002 and February 2004, and written informed consent was obtained from each individual before proceeding with genetic analysis. All clinical data, sequencing results and genotyping results were anonymous, and the study protocol was approved by the Ethical Review Committee of the National Cardiovascular Center.

From the 3655 participants of the Suita study, we selected the 96 individuals who showed the lowest levels of LDL-C to form the low LDL-C group. After analysis, it became evident that 18 of these individuals were under antihypercholesterolemia treatment; these subjects were excluded, leaving a total of 78 individuals in this group. Additionally, there were 498 individuals who were under antihypercholesterolemia medication in the total population. From this treated population we selected the 96 individuals with the highest LDL-C levels to form the treatment group. Sixteen individuals in the treatment group had ischemic heart diseases. From the untreated 3139 individuals, we selected the 96 showing the highest LDL-C levels to form the high LDL-C group.

2.2. Lipid measurements

Total serum cholesterol, triglycerides and high-density lipoprotein cholesterol (HDL-C) levels were measured with an autoanalyzer (Toshiba TBA-80; Toshiba, Tokyo, Japan) using a fasting blood sample. LDL-C levels were calculated

using the Friedewald formula [20]. Individuals with triglyceride levels higher than 400 mg/dl were omitted.

2.3. DNA analysis

Genomic DNA was isolated from blood leukocytes using an NA-3000 nucleic acid isolation system (Kurabo, Osaka, Japan). The proximal promoter (−380 bp upstream) and all coding regions (including 10 bases of each exon–intron boundary sequence) were directly sequenced. Descriptions of the primers used for amplifying and sequencing fragments are given in the electronic appendix. Polymerase chain reaction (PCR) was performed with 10 ng of genomic DNA as the template in a 10-μl reaction mixture using a HotStar Taq Master Mix Kit (Qiagen, Hilden, Germany) as follows: *Taq* polymerase was activated at 95 °C for 15 min, followed by 40 cycles of denaturing at 95 °C for 30 s, annealing at 65 °C for 30 s, and extension at 72 °C for 30 s. The PCR products were then treated with shrimp alkaline phosphatase and exonuclease I (PCR product pre-sequencing kit; USB Corporation, Cleveland, OH, USA), and used as templates for direct single-pass sequencing with a BigDye Terminator v3.1 cycle sequencing ready reaction kit (Applied Biosystems, Foster City, CA, USA). The reaction products were purified with Sephadex G-50 (Amersham Biosciences, Uppsala, Sweden) and analyzed on an ABI PRISM 3700 DNA analyzer (Applied Biosystems). The acquired sequences were examined for the presence of variants using the Sequencher software (Gene Codes Corporation, Ann Arbor, MI, USA) followed by visual inspection. Each detected variant was confirmed by repeat sequencing from the opposite direction. The A of the initiating ATG codon is denoted as nucleotide +1, and the nucleotide sequence (GenBank accession no: NT032977) was used as the reference sequence.

2.4. Statistical analysis

Lipid levels and ages were compared by Student's *t*-test. Deviations in the distributions of the genetic variants in the low LDL-C and high LDL-C/treatment groups were assessed by Fisher's two-sided exact test using the version 4 SNPAlyze statistical software (Dynacom Co. Ltd., Mobara, Japan). Pairwise linkage disequilibrium (LD) between two polymorphisms was evaluated by r^2 using SNPAlyze version 4.

3. Results

3.1. Participant characteristics

The characteristics of the individuals in the low LDL-C, high LDL-C and treatment groups, and those of the treated and untreated individuals in the total population, are shown in Table 1. The LDL-C levels of the individuals in the low LDL-C group ranged from 29.2 to 88.0 mg/dl (mean ± S.D., 70.3 ± 13.2 mg/dl), and those in the high LDL-C group ranged from 169.8 to 300.8 mg/dl (mean ± S.D., 196.7 ± 19.2 mg/dl). The LDL-C levels of

Table 1
Characteristics of individuals in the low LDL-C, high LDL-C and treatment groups, and of the total population of the Suita study

	Low LDL-C group	High LDL-C group	Treatment* group	Total population	
				Treated*	Untreated
No. (men/women)	78 (35/43)	96 (36/60)	96 (23/73)	516 (171/345)	3139 (1538/1601)
Age					
Range	39–88	40–85	44–85	44–90	35–93
Mean ± S.D.	64.2 ± 12.1	64.3 ± 10.2	69.2 ± 7.9$^{+,++}$	69.9 ± 8.5	63.9 ± 11.5
LDL-C (mg/dl)					
Range	29.2–88.0	169.8–300.8	148.0–204.4	52.4–204.4	29.2–300.8
Mean ± S.D.	70.3 ± 13.2	196.7 ± 19.2**	165.1 ± 13.4$^{+,++}$	126.9 ± 27.0	126.8 ± 29.8
Total cholesterol (mg/dl)					
Range	88–223	229–380	208–318	124–350	87–380
Mean ± S.D.	156.6 ± 22.9	277.4 ± 23.5**	246.3 ± 20.1$^{+,++}$	210.4 ± 30.1	207.6 ± 32.9
Triglycerides (mg/dl)					
Range	21–396	40–230	49–289	33–435	18–1868
Mean ± S.D.	100.7 ± 71.4	113.3 ± 43.6	118.4 ± 48.1	119.5 ± 61.9	106.2 ± 76.2
HDL-C (mg/dl)					
Range	27–125	32–106	31–91	26–108	22–140
Mean ± S.D.	66.1 ± 20.6	58.0 ± 12.4**	57.5 ± 12.6^{+}	59.4 ± 14.6	60.3 ± 15.8

LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol.

* Treatment refers to antihypercholesterolemia medication. The lipid values of the individuals in the treatment group were taken under treatment. *P* values were obtained by Student's *t*-test.

** $P < 0.005$, low LDL-C group vs. high LDL-C group.

$^{+}$ $P < 0.005$, low LDL-C group vs. treatment group.

$^{++}$ $P < 0.005$, high LDL-C group vs. treatment group.

the subjects in the treatment group ranged from 148.0 to 204.4 mg/dl (mean ± S.D., 165.1 ± 13.4 mg/dl), which were lower than those of the high LDL-C group but significantly higher than those of the untreated population (mean ± S.D., 126.8 ± 29.8 mg/dl, P value by t-test <0.005).

Triglyceride levels did not differ significantly between the three groups and these values were also similar to those of treated and untreated individuals in the total population. HDL-C levels in the low LDL-C group were significantly higher than those of all other groups. Lower LDL-C levels in general tended to be accompanied by higher HDL-C levels, however, the reason for this remains unclear. There was no difference in age between the low and high LDL-C groups, however, ages in the treatment group were higher than those of the other groups. In the total population, the treated individuals also had higher ages than the untreated individuals. This may be because exercise or diet therapy is preferred for the treatment of hypercholesterolemia in younger patients, while lipid-lowering drug therapy is often adopted in older patients.

3.2. *PCSK9 polymorphisms found*

We sequenced 156 alleles from 78 individuals with low LDL-C levels, 192 alleles from 96 individuals with high LDL-C levels, and 192 alleles from the individuals taking antihypercholesterolemia medication, identifying a total of 33 genetic variants. A list of the genetic variants and their genotype distribution for each group are shown in 2Table 2. 'Allele 1' refers to the allele shown in the GenBank reference (accession no. NT032977), and 'Allele 2' is the variant allele. An LD defined by an r^2 value greater than 0.5 is indicated in the LD column; there are four LD groups (a–d). The minor allele frequencies of the variants in groups a (−64C > T, L21-22ins, and A53V) and b (c.658 − 7C > T and c.799 + 3A > G) were greater than 0.1, indicating that these variants are relatively common. With respect to group d, the two missense mutations, A514T and V624M, were found in the same individual. The r^2 values in groups a and c were both 1.00, indicating tight LD, and the value in group b was 0.83.

Fourteen of the 33 detected variants have previously been reported (see the columns for dbSNP ID and the references in Table 2), leaving 19 novel variants. Among 14 of the variants that have been already reported, five were found to date only in the Japanese population [14], thus, a total of 24 variants are specific to the Japanese population. The genetic variants that are specific to the Japanese population are shown in boldface type in Tables 2 and 3 and Fig. 1. Twenty of the 33 variants were nonsynonymous and resulted in an amino acid change, including one nonsense mutation (W428X) in exon 8 and one trinucleotide insertion resulting in the addition of an extra leucine in a leucine repeat (L21-22ins) in exon 1. The others were either synonymous variants in coding regions or variants in untranslated regions. The sites of the nonsynonymous mutations found in the present study are shown in Fig. 1; they were distributed along the entire structure of the *PCSK9* gene.

3.3. *Contribution of the PCSK9 gene variants to plasma LDL-C levels*

Statistical analysis was carried out in order to determine whether sequence variants in the *PCSK9* gene affect plasma LDL-C levels. Deviations in the distribution of each genetic variant in either the low LDL-C group or the high LDL-C/treatment group were assessed by Fisher's two-sided exact test. Of the tested 33 genetic variants, only a single missense mutation, R93C, was found to be significantly associated with the low LDL-C group ($P = 0.003$).

Next, we compared the numbers of individuals with nonsynonymous mutations in the low LDL-C and high LDL-C/treatment groups (Table 3), since nonsynonymous mutations are presumed to have stronger and more direct effects than synonymous or noncoding sequence variants. The allele frequencies of R93C in the low LDL-C group and in the high LDL-C/treatment group were 0.051 and 0.008, respectively (Table 3). Although the other nonsynonymous mutations did not show significant statistical values in distribution, several nonsynonymous mutations were found only in the low LDL-C group (Q219E, A239D, W428X, G452D and S668R) or only in the high LDL-C/treatment group (V4I, E32K, E54A, R104C, A514T and V624M). The sequence analysis on the *LDLR* gene was performed in the individuals who had the *PCSK9* mutations found only in the high/treatment group, confirming that there was no *LDLR* mutation in these individuals. L21-22ins, A53V, A68T, G263S, I424V, V474I, V644I and G670E were found in both the low LDL-C and high LDL-C/treatment groups.

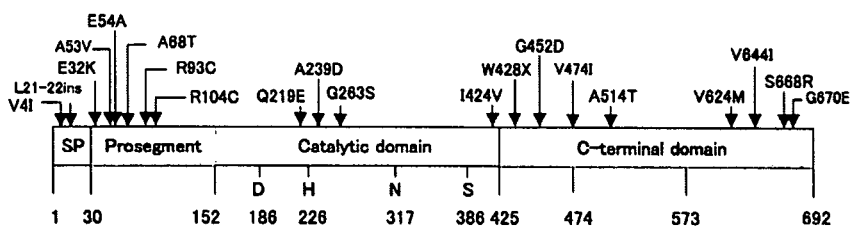

Fig. 1. Sites of the nonsynonymous mutations found in the present study. SP indicates signal peptide. The positions of the canonical aspartic acid (D), histidine (H), serine (S) catalytic triad, and oxyanion-hole asparagine (N) [3], are shown. The amino acid numbers are indicated at the bottom, and the sites of mutations found in the present study are indicated by arrows. Mutations shown in boldface type are those found only in the Japanese population.

Table 2
Genetic variants found and their genotypic distribution in each group

SNPs (allele 1 > allele 2)	LD (r^2 > 0.5)	Region	Amino acid change	No. of subjects									Allele frequency		Flanking sequences	dbSNP ID	Reference
				Low LDL-C group			High LDL-C group			Treatment group							
				Allele 1 Homo	Hetero	Allele 2 Homo	Allele 1 Homo	Hetero	Allele 2 Homo	Allele 1 Homo	Hetero	Allele 2 Homo	Allele 1	Allele 2			
−253G>A		**Exon1**		77	1	0	95	1	0	96	0	0	0.996	0.004	GCGGGCGCC[G/A]CCGTTCAGT		
−64C>T	a	**Exon 1**		63	15	0	76	18	2	72	22	2	0.883	0.117	AGGCGCCGC[C/T]GGCGTGGAC		[9,11,14,15]
c.10G>A		**Exon 1**	**V4I**	78	0	0	95	1	0	90	6	0	0.987	0.013	ATGGGCACC[G/A]TCAGCTCCA		[14]
c.63–64insCTG	a	**Exon 1**	**L21-22ins**	63	15	0	76	18	2	72	22	2	0.883	0.117	CTGCTGCTG[–/CTG]CTCCTGGGT		[9,11,14,15]
c.94G>A		**Exon1**	**E32K**	78	0	0	91	5	0	93	3	0	0.985	0.015	CGTGCGCAG[G/A]AGGACGAGG		
c.158C>T	a	**Exon 1**	**A53V**	63	15	0	76	18	2	72	22	2	0.883	0.117	ACGGCCTGG[C/T]CGAAGCACC	rs11583680	[9,11,14]
c.161A>C		**Exon 1**	**E54A**	78	0	0	96	0	0	95	1	0	0.998	0.002	GCCTGGCCG[A/C]AGCACCCGA		
c.202G>A		**Exon 1**	**A68T**	77	1	0	95	1	0	96	0	0	0.996	0.004	CACCGCTGC[G/A]CCAAGGTGC		
c.277C>T		**Exon 2**	**R93C**	70	8	0	95	1	0	94	2	0	0.980	0.020	CAGTCAGAG[C/T]GCACTGCCC		
c.310C>T		**Exon 2**	**R104C**	78	0	0	95	1	0	96	0	0	0.998	0.002	CAGGCTGCC[C/T]GCCGGGGAT		
c.336G>A		**Exon 2**	**L112L**	74	4	0	85	10	1	90	5	1	0.957	0.043	CAAGATCCT[G/A]CATGTCTTC		[14]
c.420C>T		**Exon 3**	**V140V**	78	0	0	96	0	0	95	1	0	0.998	0.002	GCCCCATGT[C/T]GACTACATC		
c.655C>G		**Exon 4**	**Q219E**	77	1	0	96	0	0	96	0	0	0.998	0.002	TTCCACAGA[C/G]AGGTAAGCA		
o.657+9G>A		**Intron 4**		73	5	0	91	5	0	93	3	0	0.976	0.024	GGTAAGCAC[G/A]GCCGTCTGA	rs11800243	[9,11,15]
o.658−7C>T	b	**Intron 4**		47	28	3	54	34	8	48	44	4	0.748	0.252	TGTGTTCGT[C/T]GAGCAGGCC	rs2483205	[9,11,15]
c.716C>A		**Exon 5**	**A239D**	77	1	0	96	0	0	96	0	0	0.998	0.002	GCCGGGATG[C/A]CGGCGTGGC		
c.787G>A		**Exon 5**	**G263S**	77	1	0	94	2	0	94	2	0	0.991	0.009	ACGGTTAGC[G/A]GCACCCTCA		
o.799+3A>G	b	**Intron 5**		42	32	4	48	39	9	46	44	6	0.715	0.285	CTCATAGGT[A/G]AGTGATGGC	rs2495477	[9,11,15]
c.993C>T		**Exon 6**	**P331P**	76	1	1	95	0	1	94	2	0	0.987	0.013	CTCAGCTCC[C/T]GAGGTAGGT		[14]
c.1217C>T		**Exon 8**	**A409A**	78	0	0	96	0	0	95	1	0	0.998	0.002	CACCCTGGC[C/T]GAGTTGAGG		
c.1270A>G		**Exon 8**	**I424V**	75	3	0	95	1	0	96	0	0	0.993	0.007	AAAGATGTC[A/G]TCAATGAGG		[14]
c.1284G>A		**Exon 8**	**W428X**	77	1	0	96	0	0	96	0	0	0.998	0.002	TGAGGCCTG[G/A]TTCCCTGAG		
c.1355G>A		**Exon 9**	**G452D**	77	1	0	96	0	0	96	0	0	0.998	0.002	TTTTTGCAG[G/A]TTGGCAGCT		
o.1380A>G	c	**Exon 9**	**V460V**	0	6	72	0	5	91	0	6	90	0.031	0.969	CAGGACTGT[A/G]TGGTCAGCA	rs540796	[9,11,14,15]
o.1420G>A	c	**Exon 9**	**V474I**	0	6	72	0	5	91	0	6	90	0.031	0.969	GCCACAGCC[G/A]TCGCCCGCT	rs562556	[9,11,14,15]
c.1540G>A	d	**Exon 10**	**A514T**	78	0	0	96	0	0	95	1	0	0.998	0.002	GCCCACAAC[G/A]CTTTTGGGG		
c.1853+6G>A		**Intron 11**		78	0	0	96	0	0	95	1	0	0.998	0.002	GCAGGTGAA[G/A]AGGCCCGTG		
c.1870G>A	d	**Exon 12**	**V624M**	78	0	0	96	0	0	95	1	0	0.998	0.002	CAGGTGACC[G/A]TGGCCTGCG		
c.1878C>T		**Exon 12**	**C626C**	78	0	0	95	1	0	96	0	0	0.998	0.002	CGTGGCCTG[C/T]GAGGAGGGC		
c.1930G>A		**Exon 12**	**V644I**	77	1	0	94	2	0	96	0	0	0.994	0.006	ACCTCCCAC[G/A]TCCTGGGGG		[14]
c.1947C>T		**Exon 12**	**A649A**	78	0	0	96	0	0	95	1	0	0.998	0.002	GGCCTACGC[C/T]GTAGACAAC		
c.2004C>A		**Exon 12**	**S668R**	77	1	0	96	0	0	96	0	0	0.998	0.002	CAGCACCAG[C/A]GAAGGGGCC		
o.2009G>A		**Exon 12**	**G670E**	0	6	72	0	13	83	0	6	90	0.046	0.954	CCAGCGAAG[G/A]GGCCGTGAC	rs505151	[9,11,15]

The mutations found only in Japanese are shown in boldface type. SNP, single nucleotide polymorphism; LD, linkage disequilibrium; homo, homozygote; hetero, heterozygote.

Table 3
No. of individuals with nonsynonymous mutations in the low LDL-C and high LDL-C/treatment groups

Amino acid change	Domain	No. of subjects			Allele frequency		LDL-C (mean ± S.D.) (mg/dl)	Identity with rodents
		Low LDL-C group (n = 78)	High LDL-C group (n = 96)	Treatment group (n = 96)	Low LDL-C group (n = 78)	High LDL-C/treatment group (n = 192)		
Mutation significantly associated with the low LDL-C group								
R93C	**Prosegment**	**8**	**1**	**2**	**0.051***	**0.008***	**96.2 ± 43.0**	**N**
Mutations found only in the low LDL-C group								
Q219E	**Catalytic**	**1**	**0**	**0**	**0.006**	**0.000**	**83.6**	**Y**
A239D	**Catalytic**	**1**	**0**	**0**	**0.006**	**0.000**	**49.8**	**Y**
W428X	**C-terminal**	**1**	**0**	**0**	**0.006**	**0.000**	**70.4**	**Y**
G452D	**C-terminal**	**1**	**0**	**0**	**0.006**	**0.000**	**73.6**	**Y**
S668R	**C-terminal**	**1**	**0**	**0**	**0.006**	**0.000**	**87.6**	**Y**
Mutations found only in the high LDL-C/treatment group								
V4I	**Signal peptide**	**0**	**1**	**6**	**0.000**	**0.018**	**168.5 ± 14.4**	**N**
E32K	**Prosegment**	**0**	**5**	**3**	**0.000**	**0.021**	**167.5 ± 10.4**	**N**
E54A	**Prosegment**	**0**	**0**	**1**	**0.000**	**0.003**	**153.0**	**N**
R104C	**Prosegment**	**0**	**1**	**0**	**0.000**	**0.003**	**192.2**	**Y**
A514T	**C-terminal**	**0**	**0**	**1**	**0.000**	**0.003**	**192.2**	**Y**
V624M	**C-terminal**	**0**	**0**	**1**	**0.000**	**0.003**	**192.2**	**Y**
Mutations found in both groups								
L21-22ins	Signal peptide	15	18,2**	22,2+	0.096	0.125	146.3 ± 50.3++	N
A53V	Prosegment	15	18,2**	22,2+	0.096	0.125	146.3 ± 50.3++	N
A68T	**Prosegment**	**1**	**1**	**0**	**0.006**	**0.003**	**129.3 ± 89.2**	**N**
G263S	**Catalytic**	**1**	**2**	**2**	**0.006**	**0.010**	**172.4 ± 52.8**	**Y**
I424V	**Catalytic**	**3**	**1**	**0**	**0.019**	**0.003**	**104.0 ± 52.6**	**Y**
V474I	C-terminal	6	5	6	0.038	0.029	141.2 ± 53.6	N
V644I	**C-terminal**	**1**	**2**	**0**	**0.006**	**0.005**	**145.9 ± 50.4**	**N**
G670E	C-terminal	6	13	6	0.038	0.049	158.8 ± 55.2	Y

Catalytic, catalytic domain; C-terminal, C-terminal domain; *P-value by Fisher's exact test was 0.003; N, amino acid residue in human wild-type has no identity with those of mouse and rat; Y, amino acid in human wild-type has identity with those of mouse and rat; **18 heterozygotes and 2 homozygotes; +22 heterozygotes and 2 homozygotes; ++mean ± S.D. was calculated from the values of the heterozygotes. The mutations specific to Japanese are shown in boldface type.